CrossMark

Collaboration facilitates abstract category learning

J. Elizabeth Richey [1,2] · Timothy J. Nokes-Malach [1] · Kara Cohen [1]

Abstract

We examined the effects of collaboration (dyads vs. individuals) and category structure (coherent vs. incoherent) on learning and transfer. Working in dyads or individually, participants classified examples from either an abstract coherent category, the features of which are not fixed but relate in a meaningful way, or an incoherent category, the features of which do not relate meaningfully. All participants were then tested individually. We hypothesized that dyads would benefit more from classifying the coherent category structure because past work has shown that collaboration is more beneficial for tasks that build on shared prior knowledge and provide opportunities for explanation and abstraction. Results showed that dyads improved more than individuals during the classification task regardless of category coherence, but learning in a dyad improved inference-test performance only for participants who learned coherent categories. Although participants in the coherent categories performed better on a transfer test, there was no effect of collaboration.

Keywords Categories · Learning · Collaboration · Metacognition

Concepts and categories are critical to many cognitive functions including classification, inference generation, explanation, problem-solving, and communication (Murphy, 2002; Ross, Taylor, Middleton, & Nokes, 2008). Although most real-world category learning happens in social contexts, the vast majority of laboratory research has focused on how individuals acquire categories on their own (Kobayashi, 1994; Miyake, 2008; van Boxtel, van der Linden, & Kanselaar, 2000). Very little work has examined the role of collaboration, or working with others, in category learning, with the exception of research on learning concepts through referential communication (Markman & Makin, 1998; Voiklis & Corter, 2012). Examining how individuals learn about novel categories through collaboration may be critical to understanding how people acquire categories more generally.

Separately, there is a large research literature on the impact of collaboration on problem-solving and memory in both

---
J. Elizabeth Richey is now at the Robert H. Smith School of Business, University of Maryland.

✉ J. Elizabeth Richey
  jerichey@rhsmith.umd.edu

1   Learning Research & Development Center, University of Pittsburgh, Pittsburgh, PA, USA

2   Robert H. Smith School of Business, 4109 Susquehanna Hall, 7699 Mowatt Lane, College Park, MD 20742-1815, USA

laboratory and educational contexts (Johnson & Johnson, 1985; Rajaram & Pereira-Pasarin, 2010; Springer, Stanne, & Donovan, 1999). Much of this work has explored the factors that impact collaborative success and failure and its relation to individual performance and learning. However, this research has not examined the impact of collaboration on a classic category-learning task. In the current work, we bring together these two rich literatures and examine the effect of collaboration on category learning. Do collaborators learn more accurately or efficiently than individuals learning alone? Do collaborators acquire more abstract category representations than those who learn individually?

In the following sections, we first review the prior work on collaboration and the factors thought to influence success and failure. We then explain why collaboration may be of particular interest for understanding category learning and review some prior research that has examined one aspect of collaboration: referential communication. We introduce abstract coherent categories (Rehder & Ross, 2001) as a particular category structure that may facilitate collaborative success and describe the present study.

Collaborative success and failure

A great deal of learning occurs through interactions with other people in school, on the job, and in social settings, making

collaboration a critical part of learning and problem-solving (Nokes-Malach, Meade, & Morrow, 2012; Sawyer, 2007; Steiner, 1972). Although collaboration is frequently deployed as a means of improving performance in academic and professional environments, prior research has shown mixed results (see Nokes-Malach, Richey, & Gadgil, 2015, for a review). Looking specifically at the impact of collaboration on conceptual learning, some research suggests that people acquire more robust concepts with fewer misconceptions when working together (Benbunan-Fich & Arbaugh, 2006; Gadgil & Nokes-Malach, 2012; Kobayashi, 1994; Lumpe & Staver, 1995; Springer et al., 1999). Other evidence, however, suggests that working with others has no lasting effect on conceptual learning compared with working alone, or that working together might actually inhibit conceptual learning (Leidner & Fuller, 1997; Pociask & Rajaram, 2014). These seemingly contradictory findings can be better understood through a brief review of some of the factors thought to support or inhibit collaborative learning.

Some research on collaboration focuses on the benefits of pooling cognitive resources (Kirschner, Paas, & Kirschner, 2009, 2011), since two people have more cognitive resources between them than one person working alone and may also possess complementary knowledge (Johansson, Andersson, & Rönnberg, 2005). Other research suggests that collaboration changes learners' cognitive processes by increasing their use of explanation (Hausmann, Nokes, VanLehn, & van de Sande, 2009; Hausmann, van de Sande, & VanLehn, 2008; Okada & Simon, 1997) and generation of abstract representations (Schwartz, 1995; Shirouzu, Miyake, & Masukawa, 2002). Still other evidence shows that collaboration can improve metacognitive awareness (Larkin, 2006), increase metacognitive behaviors (Whitebread, Bingham, Grau, Pino Pasternak, & Sangster, 2007), and support error monitoring and correction (Brodbeck & Greitemeyer, 2000; Gadgil & Nokes-Malach, 2012; Hall, Dansereau, O'Donnell, & Skaggs, 1989).

As with collaborative success, the proposed explanations for collaborative inhibition span a variety of factors. The cognitive burdens of coordinating two or more sets of memories and ideas may deplete cognitive resources that could otherwise be devoted to the task (Kirschner et al., 2009; Meade, Nokes, & Morrow, 2009). Further, group members may fail to retrieve memories (Basden, Basden, Bryner, & Thomas, 1997) or fail to develop their own ideas if they are devoting cognitive resources to listening to others' ideas, a process known as production blocking (Diehl & Stroebe, 1987).

The standard of comparison for judging collaborative success or failure is also important. Frequently, two people collaborating on a memory or learning task will perform better than the average individual (Nokes-Malach et al., 2015). However, dyads and groups are often at a disadvantage when compared with nominal groups, formed by taking two

individuals who completed the task on their own and pooling their performance (Barron, 2003; cf. Laughlin, Zander, Knievel, & Tan, 2003). For example, Weldon and Bellinger (1997) found that dyads completing a memory task performed better than individuals completing the same task but worse than nominal dyads, whereas Mullen, Johnson, and Salas (1991) conducted a meta-analysis showing brainstorming groups generally produced fewer and lower-quality ideas compared with nominal groups. A nominal dyad essentially captures the odds that either of two individuals will succeed at a task, without the costs incurred through collaboration—though also without many of the beneficial mechanisms proposed above, such as an increase in explanation or metacognitive behavior. Comparing a dyad's performance to a nominal dyad is a high bar, but comparison to a nominal group provides additional information about the degree to which working together inhibits or facilitates performance beyond what two individuals are able to do on their own.

The context of the interaction and type of task appears to play a large role in determining whether collaboration improves or inhibits learning and performance. Kirschner and colleagues argued that the pooling of multiple individuals' working-memory capacities leads to learning benefits only when task demands are great enough to require more working memory than one individual can readily supply, based on evidence that collaboration is beneficial for tasks with high cognitive demand but not for similar tasks with lower cognitive demand (Kirschner et al., 2009, 2011). Tasks that build on a shared body of knowledge among collaborators also have been found to support collaboration (Meade et al., 2009), and shared prior knowledge may support the previously mentioned processes of explanation and abstraction.

Collaboration and category learning

Many of the task features thought to support collaborative facilitation—high working-memory demand, explanation, abstraction, and a shared body of knowledge—are involved in category learning. Examining collaboration in the context of category learning is important both because of the presence of these task features and because so much category learning occurs in collaboration with others. However, very little prior work has examined the effects of learning categories in collaboration with others, and all research that has done so has focused on a highly structured form of collaboration using a referential communication paradigm. Referential communication research examines how two individuals share information, often by constructing and coordinating labels for the components of the set of objects or events being discussed (Krauss & Fussell, 1991, 1996). Referential communication tasks typically involve separating the participants so that each has access to different information, and then requiring them to

use verbal communication to coordinate what they know and complete the task. Such a task can lead to the creation of a transactive memory system, which consists of new knowledge that neither possessed on their own (Wegner, 1987).

Markman and Makin (1998) argue that because communication is a primary function of categories, it is critical to understand the role of referential communication in category construction and use. The authors had participants construct models either in a referential communication task (one partner viewing instructions and the other building the model) or by themselves. After constructing the models, all participants sorted the pieces used in the task, which served as a measure of how they categorized different pieces. While there were large individual differences in how people sorted the pieces, participants who collaborated showed much more similarity in their sorting than those who did not work together. This suggests that the process of communicating during the task shaped the categories participants formed, and that those category structures remained even once participants were working alone.

Voiklis and Corter (2012) also examined referential communication and category learning. In their experiment, participants in three conditions (collaborative, individuals with instructions to talk aloud, or individuals who worked silently) were asked to classify novel alien creatures made up of five different feature dimensions. In the collaborative setting, one person was instructed to look at an image of the creature while the other person predicted its functions. Compared with individuals, participants who collaborated performed better on the classification task. The authors argued that collaboration changed participants' cognitive processes, with collaborators coordinating their attention to specific features, applying labels to those features, and drawing inferences. In other words, they argued that collaborators had to create a more elaborate conceptual structure for communicating about the exemplars and their possible classifications. Voiklis and Corter (2012) argue that this structure is not just the result of verbalizing labels for different categories and features, because participants working individually in a think-aloud condition did not perform any better than those working individually in a silent condition.

We propose extending this work in several important ways. In the prior work by Markman and Makin (1998) and Voiklis and Corter (2012), questions about collaboration focused on referential communication, which is aimed at coordinating attention, labels, and prior knowledge (Krauss & Fussell, 1991, 1996). Both experiments use a typical paradigm in referential communication research that involves separating the participants so that each has access to different information and is assigned a specific role. Consequently, collaboration in referential communication tasks is highly structured, with dialogue typically limited to describing observable features that one partner can access. These constraints direct collaborators

toward particular cognitive processes, and it is unclear whether dyads would spontaneously engage in the same cognitive processes in more typical collaborative tasks involving joint efforts but shared access to the same resources. Collaboration in academic and professional settings is not usually as highly structured as a referential communication task, so testing the effects of collaboration on category learning outside of the referential communication paradigm is important for conclusions about how collaboration affects category learning in the real world. In more natural interactions, collaborative teams may also be more likely to focus on reaching a joint decision than when forced to take structured turns, and this type of consensus-focused collaboration has been shown to lead to better learning than forced turn-taking (Harris, Barnier, & Sutton, 2012).

Further, while this work required participants to look at combinations of features, the classification tasks in both experiments involved concrete categories because classification was based on the presence or absence of diagnostic features within a category (e.g., one type of eye meant a given creature was destructive, and the other type of eye meant it was not). Voiklis and Corter (2012) found collaborators paid greater attention to individual features and the conjunction of multiple features, so a logical next step in this line of research would be to examine abstract coherent categories, which can only be learned by recognizing the relational fit among features. Examining collaboration effects on abstract category learning is also a logical extension because it creates a learning situation in which people can benefit from applying their prior knowledge—something that may be particularly beneficial for collaborative dyads (Nokes-Malach et al., 2012).

Abstract category learning

Abstract categories are determined not by the individual features of exemplars but by the *relations* among the features, meaning the presence or absence of individual features cannot be used to determine category membership (Erickson, Chin-Parker, & Ross, 2005). Compared with a concrete-category-learning task, an abstract-category-learning task should require greater use of abstract representations, which are thought to be supported by collaboration (Schwartz, 1995; Shirouzu et al., 2002).

Research examining abstract, realistic categories with complex features has revealed that *coherence*, or the utility of prior knowledge in making sense of something, is critical for making connections among category features when learning new concepts (Lin & Murphy, 1997; Patalano, Chin-Parker, & Ross, 2006; Rehder & Ross, 2001; Spalding & Ross, 2000; Wisniewski, 1995). The members of abstract, coherent categories and their features can be extremely diverse, but they must all have the same types of relations connecting their

features. For example, although a car and motorboat are both considered *vehicles*, they do not share the same concrete features. One travels on dry land and has wheels, while the other travels on water and has a smooth bottom. They are both considered exemplars of the *vehicle* category because their features have the same relations, such as the features on the bottom of the vehicle being well suited for moving across the vehicle's environment.

Many common concepts are types of abstract coherent categories, such as natural kinds, social groups, political and military scenarios, legal concepts, and societal institutions (Rehder & Ross, 2001). Prior knowledge is critical for learning abstract coherent categories, as they are only coherent in light of prior knowledge. Shared prior knowledge is shown to help support and facilitate collaboration (Nokes-Malach et al., 2012), making this an important learning task in which to explore the effects of collaboration on category learning. Although abstract category learning appears to be especially well suited to collaboration and often occurs through interactions with others, to our knowledge no prior work has examined collaboration in the domain of abstract category learning.

The present study

This research tests whether collaboration improves abstract concept learning and whether the benefits of collaboration depend on the coherence of the category structure. The abstract-category-learning task was adapted from a series of experiments reported in Rehder and Ross (2001), in which participants were instructed to differentiate between a coherent category of pollution-cleaning devices (*morkels*) and an incoherent contrast category (*krenshaws*). Morkels and krenshaws consisted of the same features, but morkels' features were combined in a way that made sense based on prior knowledge, while krenshaws' features did not make sense. For example, a morkel might operate on the surface of water, work on absorbing spilled oil, and be coated with spongy material, while a krenshaw would operate on the surface of water, work to absorb dangerous gaseous ions, and have a shovel. Participants could not rely on individual features to distinguish morkels and krenshaws; instead, they had to look at the relations among features (coherent or incoherent). Compared with a condition that saw two sets of incoherent devices (incoherent morkels and the same incoherent krenshaw contrast condition), participants in the coherent-category condition performed better when asked to classify exemplars as morkels or krenshaws, to infer missing features from the exemplars, and to transfer their knowledge to new sets of morkels and krenshaws (Rehder & Ross, 2001). Participants in the coherent condition could classify exemplars based on the abstract relations among features, whereas

participants in the incoherent condition needed to memorize the specific classification for each five-feature exemplar.

This particular set of tasks was chosen for the present study because it includes multiple features relevant to collaboration. First, constructing abstract coherent categories requires learners to apply prior knowledge, a task feature that has also been shown to improve collaborative success. Second, it places a large demand on working memory. Collaborators have greater pooled working memory than individuals working alone, but collaboration also introduces working memory costs through the need to coordinate efforts and attend to social factors. Consequently, the benefits of pooling working memory are thought to improve performance only if the task demands a large amount of working memory (Kirschner et al., 2009). Remembering 18 exemplars, each of which has five features, is a substantial memory task for an individual; even recalling the five features of the previous exemplar while processing feedback might exceed an individual's working-memory capacity. Third, the task requires abstract reasoning, as participants in the coherent condition must focus on the abstract relations among exemplar features rather than the features themselves. Prior research has suggested that collaboration might be an effective tool for increasing participants' focus on abstract representations (Schwartz, 1995; Shirouzu et al., 2002). Finally, the task of categorizing individual exemplars during the learning phase introduces a significant opportunity for monitoring one's learning and accuracy, and collaboration has been shown to improve the metacognitive ability of error detection. Since there is a test item associated with each exemplar, the task is also well suited to measuring metacognition by asking participants to rate how accurate they think they are after each test item.

This study aims to test whether collaboration improves abstract category learning compared with learning as individuals. Based on the relevant features of the abstract-category-learning task, we predict that there will be a main effect of collaboration, with participants who learned as dyads demonstrating better category learning than individuals, possibly as a result of improved working-memory capacity, abstract reasoning, and metacognitive monitoring. It is unclear whether dyads will perform better during the learning phase than their predicted potential based on nominal dyads, but we expect they will not perform worse than nominal dyads, showing that collaborative factors reducing performance at least do not outweigh collaborative benefits. Based on the role of prior knowledge specific to learning abstract coherent categories, we expect an interaction between collaboration and coherence, such that dyads show more collaborative benefits learning coherent than incoherent category structures. We also expect a main effect of collaboration on metacognitive judgments, with participants who learned as dyads making more accurate metacognitive judgments than participants who learned as individuals. Additionally, we expect to replicate Rehder and

Ross's (2001) main effect of category structure on performance, with participants in coherent-category conditions performing better than those in incoherent-category conditions across all phases of the experiments.

Method

The experiment employed a 2 (collaboration: individual or dyad) × 2 (category type: coherent or incoherent) between-subjects design with four conditions: coherent-category dyad, coherent-category individual, incoherent-category dyad, and incoherent-category individual.

Participants

One hundred and two University of Pittsburgh undergraduates participated for course credit: 26 (13 dyads) were in a coherent-category dyad condition, 25 were in a coherent-category individual condition, 26 (13 dyads) were in an incoherent-category dyad condition, and 25 were in an incoherent-category individual condition. Thirty-eight participants identified as female, 60 identified as male, and four declined to indicate their gender. Participants were randomly assigned to conditions before reporting for the experiment, and gender was not controlled for in condition assignments. Among the dyads, there were four pairs of women (two pairs in the coherent condition and two in the incoherent condition), 14 pairs of men (six pairs in the coherent condition and eight in the incoherent condition), seven mixed-gender pairs (four pairs in the coherent condition and three in the incoherent condition), and one pair that included a person who declined to indicate their gender (coherent condition).

Materials

There were three types of materials employed in this experiment: classification task, inference test, and transfer test. All materials were presented electronically using PsyScope (J. D. Cohen, MacWhinney, Flatt, & Provost, 1993) to allow precise control over task time.

Classification task Participants viewed exemplars representing artificial categories of pollution-cleaning devices with different features across five dimensions: environment, job, material, power source, and tool. Features were combined to form 27 exemplars (see Table 1; Erickson et al., 2005; Rehder & Ross, 2001; Wisniewski, 1995). Participants in coherent-category conditions were asked to classify nine exemplars from the coherent target category ("morkels") and nine exemplars from the incoherent contrast category ("krenshaws"; see Table 2). Participants in incoherent-category conditions were asked to

Table 1. Structure of coherent and incoherent morkel and incoherent krenshaw categories

Exemplar	Dimension				
	Environment	Job	Tool	Material	Power source
Coherent morkels (target, coherent condition)					
1	1*	1	1	1	1
2	1	1*	1	2	2
3	1	1	1*	3	3
4	2*	2	2	1	3
5	2	2*	2	2	1
6	2	2	2*	3	2
7	3*	3	3	1	2
8	3	3*	3	2	3
9	3	3	3*	3	1
Incoherent morkels (target, incoherent condition)					
1	1*	2	3	1	1
2	1	2*	3	2	2
3	1	2	3*	3	3
4	3*	1	2	1	3
5	3	1*	2	2	1
6	3	1	2*	3	2
7	2*	3	1	1	2
8	2	3*	1	2	3
9	2	3	1*	3	1
Incoherent krenshaws (contrast, both conditions)					
1	1*	3	2	3	3
2	1	3*	2	2	2
3	1	3	2*	1	1
4	2*	1	3	3	2
5	2	1*	3	2	1
6	2	1	3*	1	3
7	3*	2	1	3	1
8	3	2*	1	2	3
9	3	2	1*	1	2

*Features missing in the four-feature inference test

classify nine exemplars from the incoherent target category ("morkels") and nine exemplars from the incoherent contrast category ("krenshaws"). The exemplars were presented one at a time on a computer screen, and features were always listed in the order of environment, job, tool, material, and power source. Although participants were not told this, only the first three features were needed for categorization purposes.

Inference test The four-feature inference test consisted of the same exemplars presented during the learning phase for that condition, except this time missing one of the features from the environment, job, or tool dimension (see Table 1). Participants saw each four-feature exemplar once, for a total of 18 exemplars (nine from the target and nine from the

Table 2 Category feature values

Dimension	Dimension value		
	1	2	3
Environment	Operates on surface of water	Operates on land	Floats in the stratosphere
Job	Works on absorbing spilled oil	Works to gather harmful solids	Works to absorb dangerous gaseous ions
Tool	Coated with spongy material	Has a shovel	Has an electrostatic filter
Material	Made of aluminum	Made of steel	Made of copper
Power source	Powered by battery	Powered by methane	Powered by alcohol

contrast category), with three multiple-choice options for the missing feature of each exemplar.

Transfer test The transfer test consisted of 12 novel exemplars, six of which were novel coherent morkels and six of which were novel incoherent krenshaws. These exemplars had new features but included the same five dimensions used in the learning and test materials. For example, one novel coherent morkel had the following features: operates in highway tunnels (environment), works to remove carbon monoxide (job), made of tin (material), powered by diesel fuel (power source), and has large intake fans (tool). All five features were shown, and all participants saw the same 12 novel exemplars.

The transfer test was modified from Rehder and Ross's (2001) work, in which participants saw the same novel exemplars but were asked to classify them as either morkels or nonmorkels, which were the same labels used during the initial classification task in that experiment. The labeled category of "krenshaws" in the present experiment might be viewed as more exclusive than the open category of "nonmorkels," particularly to the participants in the incoherent condition who would have needed to memorize specific combinations of features to correctly categorize exemplars during the learning phase. Therefore, an "other" option was offered for participants whose definitions of morkels or krenshaws might be specifically tied to the features in the learning phase. Our analysis primarily focuses on scoring responses as correct when participants classified coherent novel exemplars as morkels and incoherent novel exemplars as krenshaws. However, we also report results when "other" was considered the correct response for all items in the incoherent condition.

Procedure

As with the materials, the procedure consisted of three phases: a classification task, an inference test, and a transfer test. The experiment was completed in one experimental session lasting between 1 1/2 and 2 hours. In all phases, the order of presentation of each exemplar was randomized, and participants could not skip ahead or look back at their previous answers.

Classification task Participants were told they would work individually or with another participant to learn about types of items. To encourage all participants to communicate verbally, they were told they would be audio recorded during the learning phase. Those assigned to the dyad condition received no further instructions about how they should work together. Chairs for dyads were placed side-by-side, facing the same computer screen. All dyads communicated verbally, with some speaking more extensively than others. Consistent with prior work that aimed to differentiate the effects of collaboration from simply verbalizing one's thoughts, we controlled for possible effects of using verbal language by instructing individuals to "think aloud" while working through the classification-learning phase (e.g., Voiklis & Corter, 2012). Many individuals required no further prompting to verbalize their thoughts, while others needed additional encouragement. Any time a participant completed three exemplar categorizations in a row without verbalizing their thoughts, they received a scripted verbal prompt from the experimenter, such as, "Remember to talk out what you are thinking as you are classifying each object" or "Remember to try and verbalize your thought process as much as possible while you are coming to a decision."

Participants were presented with exemplars one at a time on a computer and given 45 seconds per trial to classify each exemplar as a morkel or a krenshaw. A warning that time would soon expire appeared after 35 seconds, and if no response was entered after 45 seconds, the screen would advance to the next exemplar without offering feedback. If a response was entered, participants received immediate feedback that the response was correct or incorrect. The previous exemplar's features were not shown on the feedback screen, which was intended to increase the working-memory demands of the task, but participants were permitted to remain on the feedback screen and discuss their thoughts about the feedback for up to 30 seconds per trial. Participants could advance from the feedback screen to the next trial by pressing any key at any time, or they would automatically advance after 30 seconds.

Each block consisted of all 18 exemplars, and blocks were repeated four times or until the participant or dyad reached

criterion, which was correctly classifying at least 16 of the exemplars in a block. The presentation order of the exemplars was randomized across blocks.[1]

Inference test All participants completed the four-feature inference test individually. They were presented with 18 four-feature exemplars and asked to select the missing feature from three alternatives. There was no time limit for responses. After each exemplar, participants were asked to rate their confidence in their responses on a scale from 1 (*guess*) to 7 (*certain*). The exemplars were presented in random order, and no feedback was provided.

Transfer test For the transfer test, individual participants were asked to classify novel exemplars as morkels, krenshaws, or "other." There was no time limit for responses. After each trial, participants were asked to rate their confidence in their answers on a scale from 1 (*guess*) to 7 (*certain*). The exemplars were presented in random order and no feedback was provided.

Results

Results are divided into three sections: initial classification learning, inference test, and transfer test. Across all measures, we assessed whether collaboration and category type affected performance. For the classification-learning task, we compared the performance of observed dyads and individuals by examining the number of blocks to reach criterion, improvement across trials, and the average time spent per trial in the learning phase. We also compared individuals, dyads, and nominal dyads on the probability of reaching criterion during the learning phase. To calculate predicted probability for nominal dyads, we followed a "truth wins" combination procedure that sums the probabilities that either or both individuals together could reach criterion. This procedure for predicting nominal group performance has been used in a number of studies examining the effects of collaboration on problem-solving and was determined to be the best approach for estimating nominal group performance given the categorical nature of the learning task (reaching criterion or not) and the possible performance measures available (Lorge &

Solomon, 1955; Nokes-Malach et al., 2012; Schwartz, 1995, Shirouzu et al., 2002).

For the inference and transfer tests, we examined individual participants' accuracy and confidence in their answers. Because all participants completed the inference and transfer tests as individuals, we did not compare nominal groups against those who had learned as dyads and those who learned as individuals.

We set the alpha level at .05 for main effects, interactions, and planned comparisons, and report marginal effects for *p*-values between .05 and .10 (Keppel & Wickens, 2004). We report effect sizes (Cohen's *d* or partial eta squared, η_p^2) for all main effects, interactions, and planned comparisons, and we interpret effects as small when $\eta_p^2 < .06$ or $d < .2$, medium when $.06 < \eta_p^2 < .14$ or $.3 < d < .8$, and large when $\eta_p^2 > .14$ or $d > .8$ (see Cohen, 1988; Olejnik & Algina, 2000).

Classification learning We examined the number of blocks required to reach the criterion of making no more than two mistakes in a block. If participants did not reach criterion by the end of four blocks, the learning phase ended. Table 3 indicates the percentage of dyads and individuals by condition that reached criterion during each block. To calculate predicted probability for nominal dyads, we summed the probability that one individual, the other individual, or both individuals in a nominal dyad would reach criterion. For example, in the coherent-category condition, individuals had a 16% probability of reaching criterion. The probability that a coherent-category nominal dyad with Individuals A and B would reach criterion was calculated as sum of the probability that Individual A but not B reached criterion (.16 × .84), the probability that Individual B but not A would reach criterion (.84 × .16), and the probability that both Individuals A and B would reach criterion (.16 × .16; Lorge & Solomon, 1955; Schwartz, 1995).

Table 3. Number of participants by condition reaching criterion during each block and probability of reaching criterion during classification learning phase

Block						
Condition	1	2	3	4	N/A	Probability
Coherent category						
Individual	0	1	1	2	21	.16
Observed dyad	0	1	4	0	8	.38
Nominal dyad						.29
Incoherent category						
Individual	0	0	0	0	25	0
Observed dyad	0	0	1	1	11	.15
Nominal dyad						0

---

[1] Due to an error in the program, participants in the incoherent-category conditions generally saw one less exemplar per block (*M* = 17.05, *SD* = .10). This potentially put them at a disadvantage for reaching criterion (16 correct responses in a block); however, a careful review of the data showed that no participant in the incoherent-category conditions came within one correct item of reaching criterion, meaning an additional trial would not have created the opportunity to reach criterion. All time and performance data are averaged based on the number of trials completed, which accounts for the variance in number of blocks participants saw based on whether they reached criterion, as well as the variance in number of trials per block based on the program error.

Within the coherent-category condition, the rate of reaching criterion for observed dyads was marginally different from the rate for individuals, $z = 1.54$, $p = .061$, and not significantly different from the predicted rate for nominal dyads, $z = 0.49$, $p = .31$. Within the incoherent-category condition, the rate of reaching criterion for observed dyads was significantly greater than the rate for individuals, $z = 2.01$, $p = .022$, and marginally greater than the rate for predicted nominal dyads, $z = 1.47$, $p = .071$.

We conducted additional analyses comparing observed dyads' and individuals' performance on the classification task. A 2×2 between-subjects analysis of variance (ANOVA) was performed to investigate effects of collaboration and category type on number of blocks to criterion. Similar to Rehder and Ross (2001), we adopted the conservative estimate that participants who never reached criterion would have done so on the next (fifth) block. Analyses revealed a medium effect of collaboration, $F(1, 72) = 5.38$, $p = .023$, $\eta_p^2 = .070$, and a medium effect of category type, $F(1, 72) = 6.96$, $p = .010$, $\eta_p^2 = .088$, for the total number of blocks to reach criterion, with dyads requiring fewer blocks than individuals and participants in coherent-category conditions requiring fewer blocks than participants in incoherent-category conditions. There was no interaction, $F(1, 72) = 0.69$, $p = .41$, $\eta_p^2 = .010$.

We also examined participants' improvement from the first block to the final block they completed, which was calculated by subtracting first-block accuracy from final-block accuracy. A 2×2 between-subjects ANOVA revealed a small effect of collaboration, $F(1, 72) = 4.38$, $p = .040$, $\eta_p^2 = .057$, with dyads demonstrating greater improvement than individuals (see Fig. 1). There was no effect of category type, $F(1, 72) = .87$, $p = .35$, $\eta_p^2 = .012$, and no interaction, $F(1, 72) = .80$, $p = .38$, $\eta_p^2 = .011$.

Looking at decision times, there was a medium effect of collaboration, $F(1, 72) = 4.85$, $p = .031$, $\eta_p^2 = .063$, with dyads spending more time than individuals. There was no effect of category type, $F(1, 73) = .33$, $p = .157$, $\eta_p^2 = .005$, and no interaction, $F(1, 72) = 1.41$, $p = .124$, $\eta_p^2 = .019$. For feedback times, there was no effect of collaboration,

$F(1, 72) = 2.09$, $p = .15$, $\eta_p^2 = .028$, and no effect of category type, $F(1, 72) = .90$, $p = .35$, $\eta_p^2 = .012$. However, there was a small interaction, $F(1, 72) = 5.27$, $p = .025$, $\eta_p^2 = .068$. To interpret the interaction, we conducted planned pairwise comparisons. There was a large effect of collaboration in coherent-category conditions, $F(1, 36) = 8.08$, $p = .007$, $\eta_p^2 = .18$, with dyads taking more time to study feedback than individuals. There were no differences between dyads and individuals in the incoherent-category condition, $F(1, 36) = .32$, $p = .58$, $\eta_p^2 = .009$ (see Fig. 2).

Additional analyses were done to examine the relation between time spent on each trial and accuracy. For dyads, there was no correlation between accuracy and decision time, $r(26) = .15$, $p = .48$, but there was a correlation between accuracy and feedback time, $r(26) = .50$, $p = .010$. Looking at individuals' times, there were no correlations between accuracy and decision time, $r(50) = .001$, $p = .99$, or accuracy and feedback time, $r(50) = -.17$, $p = .23$. These results suggest that dyads engaged in a different type of processing compared with individuals when viewing feedback, as their time spent viewing feedback was associated with accuracy while individuals' was not.

Finally, we examined whether participants in each condition were performing significantly better or worse than chance (set at .5, as each item had two possible responses). A one-sample, two-tailed t test indicated that among participants who learned as dyads, those in the coherent-category condition performed significantly above chance across the category learning phase, $M = .56$, $SD = .11$, $t(12) = 2.19$, $p = .049$, while participants in the incoherent-category condition did not perform significantly above or below chance, $M = .50$, $SD = .11$, $t(12) = 0.14$, $p = .89$. Among participants learning as individuals, those in the coherent-category condition performed marginally better than chance, $M = .54$, $SD = .087$, $t(24) = 2.03$, $p = .054$, and those in the incoherent-category condition performed marginally below chance, $M = .47$, $SD = .07$, $t(24) = -1.89$, $p = .071$.

Inference test A 2×2 between-subjects ANOVA assessing condition effects on inference-test performance revealed a
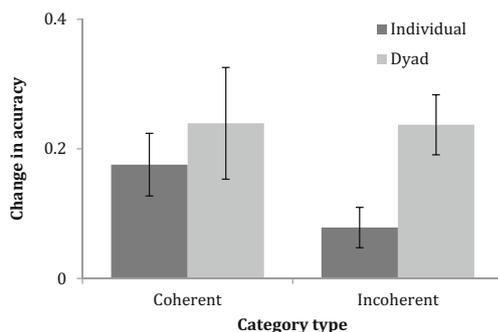
Fig. 1 Learning condition effect on change in accuracy from first to final learning block. Bars represent standard errors
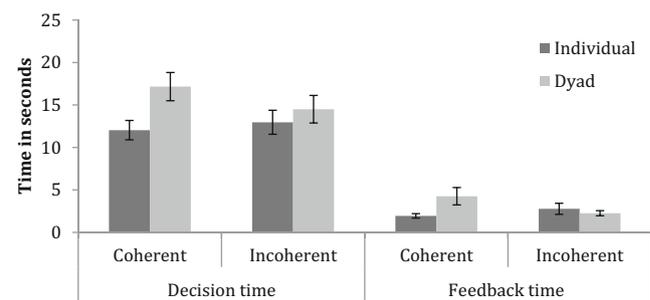

Fig. 2 Learning condition effect on average time per trial during the learning phase. Bars represent standard errors

large effect of category type, $F(1, 98) = 90.63$, $p < .001$, $\eta \eta_p^2 = .48$, with participants in coherent-category conditions performing better than participants in incoherent-category conditions (see Fig. 3). There was no effect of collaboration, $F(1, 98) = .47$, $p = .50$, $\eta_p^2 = .005$. However, there was a medium interaction, $F(1, 98) = 8.01$, $p = .006$, $\eta_p^2 = .076$. To interpret the interaction, we conducted planned pairwise comparisons. Results showed a medium effect of collaboration in coherent-category conditions, $F(1, 49) = 4.84$, $p = .032$, $\eta_p^2 = .090$, with dyads performing better than individuals. There was a marginal effect of collaboration in incoherent-category conditions, $F(1, 49) = 3.17$, $p = .081$, $\eta_p^2 = .061$, with individuals performing better than dyads.

We again examined whether participants in each condition were performing significantly better or worse than chance (set at .33, as each item had three possible responses). A one-sample, two-tailed t test indicated that among participants who learned as dyads, those in the coherent-category condition performed significantly above chance on the inference test, $M = .58$, $SD = .20$, $t(25) = 6.60$, $p < .001$, while participants in the incoherent-category condition performed significantly below chance, $M = .17$, $SD = .13$, $t(25) = -6.40$, $p < .001$. Likewise, among participants who learned as individuals, those in the coherent-category condition performed significantly above chance, $M = .47$, $SD = .18$, $t(24) = 3.75$, $p = .001$, and those in the incoherent-category condition performed significantly below chance, $M = .24$, $SD = .16$, $t(24) = -2.68$, $p = .013$.

Metacognitive judgments of the inference test Using participants' confidence ratings and accuracy on inference test trials, we examined the effect of condition on students' metacognitive judgments by calculating a discrimination score for each participant (Schraw, 2009). Discrimination scores reflect learners' abilities to differentiate between correct and incorrect responses when rating their confidence (Mengelkamp & Bannert, 2010). A positive value indicates that the learner gave higher confidence ratings for correct responses compared with incorrect responses, a negative value indicates higher confidence ratings for incorrect responses
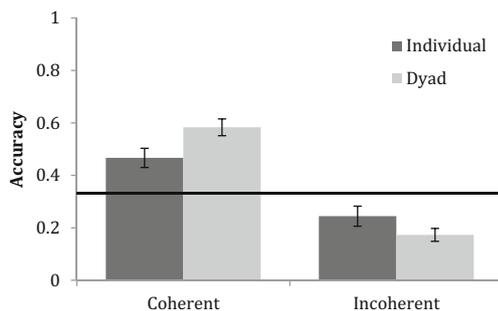
compared with correct responses, and a zero indicates no relation between the two (i.e., the learner was equally confident in correct and incorrect responses).

A 2 × 2 between-subjects ANOVA assessing condition effects on discrimination revealed a large effect of category type, $F(1, 98) = 75.29$, $p < .001$, $\eta_p^2 = .43$, with participants in incoherent-category conditions expressing greater confidence in incorrect answers compared with correct answers (i.e., negative discrimination scores) and participants in coherent-category conditions expressing greater confidence in correct answers compared with incorrect answers (i.e., positive discrimination scores; see Fig. 4). There was no effect of collaboration, $F(1, 98) = 1.25$, $p = .27$, $\eta_p^2 = .013$. However, there was a marginal interaction, $F(1, 98) = 3.25$, $p = .075$, $\eta_p^2 = .032$. To interpret the interaction, we conducted planned pairwise comparisons. Results showed a marginal effect of collaboration in coherent-category conditions, $F(1, 49) = 2.98$, $p = .091$, $\eta_p^2 = .057$, with participants who learned as dyads expressing greater confidence in correct answers than those who learned as individuals. There was no difference between those who learned as dyads and individuals in incoherent-category conditions, $F(1, 49) = .41$, $p = .52$, $\eta_p^2 = .008$. This provides weak evidence that working with a partner during the learning phase led to more accurate metacognitive monitoring during the inference test, but only for participants in coherent-category conditions.

Transfer test A 2 × 2 between-subjects ANOVA revealed a medium effect of category type, $F(1, 98) = 12.54$, $p = .001$, $\eta_p^2 = .11$, with participants in coherent-category conditions performing better than participants in incoherent-category conditions (see Fig. 5). There was no effect of collaboration, $F(1, 98) < .001$, $p = .99$, $\eta_p^2 = .001$, and no interaction, $F(1, 98) = .80$, $p = .37$, $\eta_p^2 = .008$.

We examined whether participants in each condition were performing significantly better or worse than chance (set at .33, as each item had three possible responses). A one-sample, two-tailed t test indicated that among participants who learned as dyads, those in the coherent-category condition performed significantly above chance
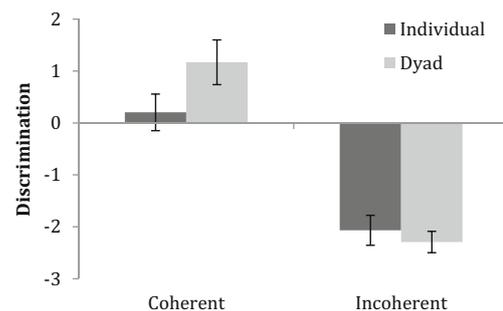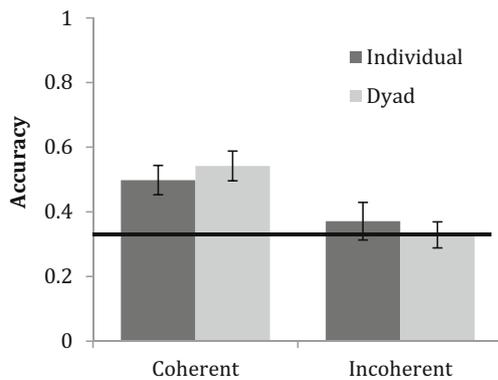

Fig. 3 Learning condition effect on inference-test accuracy, with expected performance at chance (.33) shown as a benchmark line. Bars represent standard errors

Fig. 4 Participants' metacognitive monitoring judgments of test questions as measured by discrimination (zero indicates no correlation between judgment and accuracy). Bars represent standard errors

Fig. 5 Learning condition effect on transfer test accuracy, with expected performance at chance (.33) shown as a benchmark line. Bars represent standard errors

Fig. 6 Participants' monitoring judgments of transfer questions as measured by discrimination (zero indicates no correlation between judgment and accuracy). Bars represent standard errors

across the inference test, $M = .54$, $SD = .30$, $t(25) = 3.64$, $p = .001$, while participants in the incoherent-category condition did not perform above or below chance, $M = .33$, $SD = .21$, $t(25) = -0.033$, $p = .97$. Likewise, among participants who learned as individuals, those in the coherent-category condition performed significantly better than chance, $M = .50$, $SD = .23$, $t(24) = 3.70$, $p = .001$, and those in the incoherent-category condition did not perform above or below chance, $M = .37$, $SD = .23$, $t(24) = 0.89$, $p = .38$.

If a participant in the incoherent condition memorized specific feature pairs to correspond with the labels "morkel" and "krenshaw," then all novel items on the transfer test would be categorized as "other." We recoded scores using "other" as the correct response for participants in the incoherent condition. A 2×2 between-subjects ANOVA of the recoded data revealed a small effect of category type, $F(1, 98) = 5.92$, $p = .017$, $\eta_p^2 = .057$, with participants in coherent-category conditions ($M = .52$, $SD = .26$) performing better than participants in incoherent-category conditions ($M = .42$, $SD = .11$). There was no effect of collaboration, $F(1, 98) = .32$, $p = .58$, $\eta_p^2 = .003$, and no interaction, $F(1, 98) = .28$, $p = .60$, $\eta_p^2 = .003$.

Metacognitive judgments of the transfer test We calculated discrimination scores based on participants' confidence ratings during the transfer test using the morkel-coherent, krenshaw-incoherent coding. A 2×2 between-subjects ANOVA assessing condition effects on discrimination revealed a medium effect of category type, $F(1, 98) = 8.55$, $p = .004$, $\eta_p^2 = .080$, with participants in incoherent-category conditions expressing greater confidence in incorrect answers compared with correct answers (i.e., negative discrimination scores) and participants in coherent-category conditions expressing greater confidence in correct answers compared with incorrect answers (i.e., positive discrimination scores; see Fig. 6). There was no effect of collaboration, $F(1, 98) = .050$, $p = .82$, $\eta_p^2 = .001$, and no interaction, $F(1, 98) = .75$, $p = .39$, $\eta_p^2 = .008$.
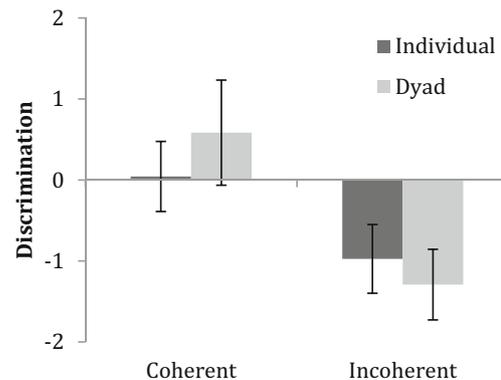
Relations between learning, test, and transfer Given the significant effects of collaboration on learning and test performance but not on transfer performance, we assessed correlations across the three measures. We conducted correlations separately for participants in each of the four conditions, since the knowledge transferring between tasks differed depending on condition (see Table 4). Specifically, in the coherent conditions, learners could transfer abstract knowledge of categories to both the inference and transfer tests or concrete knowledge of feature pairings to only the inference test. In the incoherent conditions, learners could transfer only concrete knowledge of feature pairings, and to only the inference test. In the individual conditions, learners could transfer knowledge they constructed as individuals during the learning phase. In the dyad conditions, learners could transfer the individual knowledge they constructed and the transactive knowledge co-constructed with partners, but they could not transfer knowledge held only by their partners during the learning phase. Results showed that for participants in the coherent conditions, performance during the learning phase predicted inference and transfer test performance and inference test performance predicted transfer test performance. For participants in the incoherent conditions, there were no significant relations across performance on each measure.

Discussion

This experiment brought together two robust literatures to examine the effects of collaboration on abstract category learning. Very little prior work has bridged these two literatures, even though the features of the abstract category-learning task align well with a number of collaborative mechanisms including the important role of prior knowledge, increased working-memory capacity, explanation, generation of abstract representations, and metacognitive monitoring. Although the present study cannot disentangle precisely

Table 4. Correlations between learning, inference test, and transfer performance by condition

Correlation examined	Condition			
	Dyad/ Coherent ($n = 26$)	Dyad/ Incoherent ($n = 26$)	Individual/ Coherent ($n = 25$)	Individual/ Incoherent ($n = 25$)
Learning-inference test	.53*	.21	.65*	.26
Learning-transfer test	.45*	-.045	.62*	.13
Inference test-transfer test	.58*	.033	.62*	.19

*$p < .05$

which mechanisms of collaboration might be responsible for improving performance, it suggests some have played a larger role than others. Below, we discuss implications of the results for our understanding of category learning and collaboration. We also identify several open questions and directions for future research.

Category learning

Collaboration effects Consistent with predictions, dyads performed better than individuals across multiple measures of the classification-learning task. Regardless of whether they were in the coherent-category or incoherent-category condition, dyads required fewer blocks to reach criterion. As they worked through the learning phase, dyads also improved their accuracy more than individuals. This is consistent with mechanisms thought to underpin both collaborative inhibition and facilitation: the cognitive demands of collaboration can reduce performance initially, but as partners improved their coordination with practice, the cognitive advantages of collaboration begin to emerge. Dyads did not perform worse than nominal dyads, suggesting that the benefits of collaboration were not outweighed by its costs.

There was no interaction between collaboration and coherence on any measure of classification-task performance. Learning the coherent and incoherent classification tasks could be accomplished through different mechanisms. For example, coherent classification could be accomplished either by applying prior knowledge to recognize abstract relations or by memorizing all of the exemplars. In contrast, participants in the incoherent classification conditions likely memorized exemplars because no prior knowledge could be applied. This suggests dyads in the coherent and incoherent conditions likely adopted two different but beneficial strategies. The incoherent-category condition created especially large working-memory demands, as there were no abstract relations to learn, and dyads in the incoherent condition would have needed to pool their memories. In this case, the advantage of collaboration would not carry over to dyads' performance on the individual inference test for the incoherent-category condition, when they no longer had access to the pooled memory

resources shared between collaborators. This is consistent with correlation analyses showing no relations between performance on the learning phase and test phases in the incoherent conditions.

The advantages of collaboration would carry over for coherent-condition dyads who identified the abstract relations together, as they would still have access to knowledge of abstract relations when working individually. Based on this explanation, we would expect to see an interaction between collaboration and coherence emerge on the individual inference test, with those who learned as dyads in the coherent condition benefitting from collaboration more than those in the incoherent condition. We observed this effect, supporting the idea that the collaborative advantages afforded to dyads carried over to individual performance in the coherent condition but not in the incoherent condition. In fact, participants who learned as dyads in the incoherent-category condition performed marginally worse than those who learned as individuals in the incoherent-category condition, suggesting that their reliance on pooled working memory resources during the classification-learning task was harmful to them when those pooled resources were no longer available.

An interaction was found for the amount of time participants spent discussing feedback during the learning phase. Dyads spent more time than individuals discussing feedback, but only in the coherent-category conditions. This result is also consistent with deep learning strategies, as it suggests that dyads spent longer trying to make sense of feedback when prior knowledge was relevant and the category structure included abstract relations among features. Spending more time discussing feedback was associated with greater accuracy, but only for dyads. Dyads in the incoherent condition did not spend longer on feedback than individuals in the incoherent condition. Given that participants could not view the exemplar on the feedback screen, this is more consistent with a memorization strategy that relied on having exemplar information available.

There were no effects of collaboration on the transfer test. The correlations among learning, test, and transfer were significant among participants in both individual-coherent and collaborative-coherent conditions, and both groups performed

significantly above chance. Although the effect of collaboration was weaker than category structure across learning and test phases, it is nevertheless unclear why the collaborative effect disappeared on the transfer task. Future work should examine the extent to which dyads are able to abstract and generalize materials and whether other types of transfer tasks might capture effects of collaboration.

Category type effects There was a main effect of category type on the classification task, inference test, and transfer test, with participants in coherent-category conditions performing better than participants in incoherent-category conditions. This robust effect replicates Rehder and Ross (2001) and is consistent with much past work examining the roles of prior knowledge and coherence in category learning.

Collaboration and metacognition

No main effect of collaboration was found on learners' metacognitive judgments, but there was a marginal interaction such that participants who collaborated during the learning phase demonstrated marginally more accurate discrimination between correct and incorrect responses during the inference test in the coherent-category condition. By encouraging explicit discussion of cognitive strategies and errors as partners work to coordinate efforts, collaboration has been shown to support learners' metacognitive behaviors and strategies (Kuhn, Shaw, & Felton, 1997; Larkin, 2006; Whitebread et al., 2007).

Participants in the incoherent-category conditions had negative discrimination scores on the inference and transfer tests, indicating that they were more confident in items that they answered incorrectly than in items that they answered correctly. Negative discrimination scores are regarded as a "lack of metacognitive awareness" (Schraw, 2009, p. 41), but observing this type of score is uncommon (Stankov, Kleitman, & Jackson, 2015). A possible explanation lies in incoherent-category participants' performance below chance on the inference test. This indicates more than a failure to memorize feature pairings; rather, it suggests that participants in the incoherent-category condition may have actually learned incorrect rules or feature pairings, pushing their performance below chance. Responding based on these incorrect rules or pairings would explain both why they performed worse than if they were simply guessing, and also why they were more confident in those incorrect responses than in their correct responses.

Contributions and future directions

The results have important implications for our understanding of category learning and indicate that conclusions about how people acquire categories should be reexamined in social contexts. Such an examination of collaborative influences might change our understanding of when and how people are more likely to generate abstract category representations versus exemplar-based representations. Many domains of cognitive research have incorporated socio-cognitive contexts by examining, for example, the situative perspective of learning and cognition (Greeno, 1998), social aspects of causal learning (Shafto, Goodman, & Frank, 2012), and social and cultural cognition (Tomasello, Carpenter, Call, Behne, & Moll, 2005). The results of this experiment demonstrate how applying a social lens could deepen our understanding of the conditions and mechanisms of category and concept acquisition.

The purpose of this article was not to test different mechanisms of collaboration but to test the effects of collaboration on abstract category learning. Although this work provides clear evidence of collaboration's effect on abstract category learning, it cannot clearly distinguish among multiple cognitive processes at play during collaborative learning. Future work should further test *how* collaboration improved category learning by comparing different mechanisms hypothesized to support collaborative learning. For example, one could alter the task to test specific hypotheses about which processes of collaboration might be most important. Researchers could also employ additional measures to better understand the kinds of representations collaborators form together, and how those representations change when they complete tasks on their own following a collaboration.

In addition to examining the types of representations collaborators form, it would be interesting to examine the effects of the prior knowledge participants bring to the task. For example, how would effects compare when dyads have similar types versus different types of prior knowledge, such as when one partner has more relevant prior knowledge and the other has less? This is a particularly important question because prior knowledge often differs between collaborators, such as when parents work with children or teachers collaborate with students. Finally, the improvement dyads showed over the course of the classification-learning task suggests that collaboration becomes more beneficial with practice. Future work could test this by providing dyads with some collaborative training or an opportunity to practice working together before beginning the target learning task.

Conclusion

In many learning and educational contexts, the goal of collaboration is not simply to improve performance while two or more people are working together. Although the products that groups create together may be evaluated or used, the intention is often to create new knowledge that each individual can take away after the collaborative experience is over. In these

results, learners retained the benefits of collaboration only when they were learning content with abstract relations, which has important implications for how collaborative learning tasks are structured. Results should be examined and replicated using more realistic academic content to better understand the instructional implications of this research. If the goal is to improve individuals' knowledge, then collaborative learning experiences are likely to be more fruitful when there are richer conceptual components to what is being learned or relevant prior knowledge to be applied to the collaborative learning experience.

References

Barron, B. (2003). When smart groups fail. *Journal of the Learning Sciences*, *12*(3), 307–359. https://doi.org/10.1207/S15327809JLS1203_1

Basden, B. H., Basden, D. R., Bryner, S., & Thomas, R. L. (1997). A comparison of group and individual remembering: Does collaboration disrupt retrieval strategies? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(5), 1176–1191. https://doi.org/10.1037/0278-7393.23.5.1176

Benbunan-Fich, R., & Arbaugh, J. B. (2006). Separating the effects of knowledge construction and group collaboration in learning outcomes of web-based courses. *Information & Management*, *43*(6), 778–793. https://doi.org/10.1016/j.im.2005.09.001

Brodbeck, F. C., & Greitemeyer, T. (2000). Effects of individual versus mixed individual and group experience in rule induction on group member learning and group performance. *Journal of Experimental Social Psychology*, *36*(6), 621–648. https://doi.org/10.1006/jesp.2000.1423

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd). Hillsdale: Erlbaum. https://doi.org/10.1234/12345678

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavior Research Methods, Instruments, and Computers*, *25*(2), 257–271. https://doi.org/10.3758/BF03204507

Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology*, *53*(3), 497–509. https://doi.org/10.1037/0022-3514.53.3.497

Erickson, J. E., Chin-Parker, S., & Ross, B. H. (2005). Inference and classification learning of abstract coherent categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 86–99. https://doi.org/10.1037/0278-7393.31.1.86

Gadgil, S., & Nokes-Malach, T. J. (2012). Overcoming collaborative inhibition through error correction: A classroom experiment. *Applied Cognitive Psychology*, *26*(3), 410–420. https://doi.org/10.1002/acp.1843

Greeno, J. G. (1998). The situativity of knowing, learning, and research. *American Psychologist*, *53*(1), 5–26. https://doi.org/10.1037/0003-066X.53.1.5

Hall, R., Dansereau, D., O'Donnell, A. M., & Skaggs, L. (1989). The effect of textual errors on dyadic and individual learning. *Journal of Literacy Research*, *21*(2), 127–140. https://doi.org/10.1080/10862968909547665

Harris, C. B., Barnier, A. J., & Sutton, J. (2012). Consensus collaboration enhances group and individual recall accuracy. *The Quarterly Journal of Experimental Psychology, 65*(1), 179–194.

Hausmann, R. G. M., Nokes, T. J., VanLehn, K., & van de Sande, B. (2009). Collaborative dialog while studying worked-out examples. In V. Dimitrova, R. Mizoguchi, B. Du Boulay, & A. C. Graesser (Eds.), Artificial intelligence in education (pp. 596–598). Amsterdam: IOS Press.

Hausmann, R. G. M., van de Sande, B., & VanLehn, K. (2008). Shall we explain? Augmenting learning from intelligent tutoring systems and peer collaboration. In B. P. Woolf, E. Aimeur, R. Nkambou, & S. Lajoie (Eds.), Intelligent tutoring systems (pp. 636–645). Berlin: Springer-Verlag.

Johansson, N. O., Andersson, J., & Rönnberg, J. (2005). Compensating strategies in collaborative remembering in very old couples. *Scandinavian Journal of Psychology*, *46*, 349–359. https://doi.org/10.1111/j.1467-9450.2005.00465.x

Johnson, D. W., & Johnson, R. T. (1985). The internal dynamics of cooperative learning groups. In R. Slaving, S. Sharan, S. Kagan, R. Hertz-Lazarowitz, C. Webb, & R. Schmuck (Eds.), Learning to cooperate, cooperating to learn (pp. 103–134). New York: Plenum Press.

Keppel, G., & Wickens, T. D. (2004). Design and analysis: A researcher's handbook (4th editio.). Upper Saddle River: Pearson Education International.

Kirschner, F., Paas, F. G. W. C., & Kirschner, P. A. (2009). A cognitive load approach to collaborative learning: United brains for complex tasks. *Educational Psychology Review*, *21*(1), 31–42. https://doi.org/10.1007/s10648-008-9095-2

Kirschner, F., Paas, F. G. W. C., & Kirschner, P. A. (2011). Task complexity as a driver for collaborative learning efficiency: The collective working-memory effect. *Applied Cognitive Psychology*, *25*(4), 615–624. https://doi.org/10.1002/acp.1730

Kobayashi, Y. (1994). Conceptual acquisition and change through social interaction. *Human Development*, *37*(4), 233–241. https://doi.org/10.1159/000278265

Krauss, R. M., & Fussell, S. R. (1991). Perspective-taking in communication: Representations of others' knowledge in reference. *Social Cognition*, *9*(1), 2–24. https://doi.org/10.1521/soco.1991.9.1.2

Krauss, R. M., & Fussell, S. R. (1996). Social psychological models of interpersonal communication. In E. T. Higgins & A. W. Kruglanski (Eds.), Social psychology: Handbook of basic principles (pp. 655–701). New York: Guilford Press.

Kuhn, D., Shaw, V., & Felton, M. (1997). Effects of dyadic interaction on argumentive reasoning. *Cognition and Instruction*, *15*(3), 287–315. https://doi.org/10.1207/s1532690xci1503_1

Larkin, S. (2006). Collaborative group work and individual development of metacognition in the early years. *Research in Science Education*, *36*(1/2), 7–27. https://doi.org/10.1007/s11165-006-8147-1

Laughlin, P. R., Zander, M. L., Knievel, E. M., & Tan, T. K. (2003). Groups perform better than the best individuals on letters-to-numbers problems: Informative equations and effective strategies. *Journal of Personality and Social Psychology*, *85*(4), 684–694. https://doi.org/10.1037/0022-3514.85.4.684

Leidner, D. E., & Fuller, M. (1997). Improving student learning of conceptual information: GSS supported collaborative learning vs. individual constructive learning. *Decision Support Systems*. https://doi.org/10.1016/S0167-9236(97)00004-3

Lin, E. L., & Murphy, G. L. (1997). Effects of background knowledge on object categorization and part detection. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(4), 1153–1169. https://doi.org/10.1037/0096-1523.23.4.1153

Lorge, I., & Solomon, H. (1955). Two models of group behaviour in the solution of eureka-type problems. *Psychometrika, 20,* 139–148.

Lumpe, A. T., & Staver, J. R. (1995). Peer collaboration and concept development: Learning about photosynthesis. *Journal of Research in Science Teaching*, *32*(1), 71–98. https://doi.org/10.1002/tea.3660320108

Markman, A. B., & Makin, V. S. (1998). Referential communication and category acquisition. *Journal of Experimental Psychology: General*, *127*(4), 331–354. https://doi.org/10.1037/0096-3445.127.4.331

Meade, M. L., Nokes, T. J., & Morrow, D. G. (2009). Expertise promotes facilitation on a collaborative memory task. *Memory*, *17*(1), 39–48. https://doi.org/10.1080/09658210802524240

Mengelkamp, C., & Bannert, M. (2010). Accuracy of confidence judgments: Stability and generality in the learning process and predictive validity for learning outcome. *Memory & Cognition*, *38*(4), 441–51. https://doi.org/10.3758/MC.38.4.441

Miyake, N. (2008). Conceptual change through collaboration. In S. Vosniadou (Ed.), International handbook on research in conceptual change (pp. 453–478). New York: Routledge.

Mullen, B., Johnson, C., & Salas, E. (1991). Productivity loss in brainstorming groups: A meta-analytic integration. *Basic and Applied Social Psychology*, *12*(1), 3–23. https://doi.org/10.1207/s15324834basp1201_1

Murphy, G. L. (2002). The big book of concepts. Cambridge: MIT Press.

Nokes-Malach, T. J., Meade, M. L., & Morrow, D. G. (2012). The effect of expertise on collaborative problem solving. *Thinking & Reasoning*, *18*(1), 32–58. https://doi.org/10.1080/13546783.2011.642206

Nokes-Malach, T. J., Richey, J. E., & Gadgil, S. (2015). When is it better to learn together? Insights from research on collaborative learning. *Educational Psychology Review*, *27*(4), 645–656. https://doi.org/10.1007/s10648-015-9312-8

Okada, T., & Simon, H. A. (1997). Collaborative discovery in a scientific domain. *Cognitive Science*, *21*(2), 109–146. https://doi.org/10.1207/s15516709cog2102_1

Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational Psychology*, *25*(3), 241–286. doi: https://doi.org/10.1006/ceps.2000.1040

Patalano, A. L., Chin-Parker, S., & Ross, B. H. (2006). The importance of being coherent: Category coherence, cross-classification, and reasoning. *Journal of Memory and Language*, *54*(3), 407–424. https://doi.org/10.1016/j.jml.2005.10.005

Pociask, S., & Rajaram, S. (2014). The effects of collaborative practice on statistical problem solving: Benefits and boundaries. *Journal of Applied Research in Memory and Cognition*, *3*(4), 252–260. https://doi.org/10.1016/j.jarmac.2014.06.005

Rajaram, S., & Pereira-Pasarin, L. P. (2010). Collaborative memory: Cognitive research and theory. *Perspectives on Psychological Science*, *5*(6), 649–663. https://doi.org/10.1177/1745691610388763

Rehder, B., & Ross, B. H. (2001). Abstract coherent categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(5), 1261–1275. https://doi.org/10.1037/0278-7393.27.5.1261

Ross, B. H., Taylor, E. G., Middleton, E. L., & Nokes, T. J. (2008). Concept and category learning in humans. In H. L. Roediger (Ed.), Cognitive psychology of memory (Vol. 2). Oxford: Elsevier.

Sawyer, R. K. (2007). Group genius: The creative power of collaboration. New York: Basic Books.

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning*, *4*(1), 33–45. https://doi.org/10.1007/s11409-008-9031-3

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/10.1207/s15327809jls0403_3

Shafto, P., Goodman, N. D., & Frank, M. C. (2012). Learning from others: the consequences of psychological reasoning for human learning. *Perspectives on Psychological Science*, *7*(4), 341–351. https://doi.org/10.1177/1745691612448481

Shirouzu, H., Miyake, N., & Masukawa, H. (2002). Cognitively active externalization for situated reflection. *Cognitive Science*, *26*(4), 469–501. https://doi.org/10.1016/S0364-0213(02)00066-6

Spalding, T. L., & Ross, B. H. (2000). Concept learning and feature interpretation. *Journal of Memory and Cognition*, *28*(3), 439–451. https://doi.org/10.3758/BF03198559

Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research*, *69*(1), 21–51. https://doi.org/10.3102/00346543069001021

Stankov, L., Kleitman, S., & Jackson, S. A. (2015). Measures of the trait of confidence. In G. J. Boyle, D. H. Saklofske, & G. Matthews (Eds.), Measures of personality and social psychological constructs (pp. 158–189). London: Academic Press.

Steiner, I. D. (1972). Group processes and productivity. New York: Academic Press.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*(05). https://doi.org/10.1017/S0140525X05000129

van Boxtel, C., van der Linden, J., & Kanselaar, G. (2000). Collaborative learning tasks and the elaboration of conceptual knowledge. *Learning and Instruction*, *10*(4), 311–330. https://doi.org/10.1016/S0959-4752(00)00002-5

Voiklis, J., & Corter, J. E. (2012). Conventional wisdom: negotiating conventions of reference enhances category learning. *Cognitive Science*, *36*(4), 1–28. https://doi.org/10.1111/j.1551-6709.2011.01230.x

Wegner, D. M. (1987). Transactive memory: A contemporary analysis of the group mind. In B. Mullen & G. R. Goethals (Eds.), Theories of group behavior (pp. 185–208). New York: Springer.

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

Whitebread, D., Bingham, S., Grau, V., Pino Pasternak, D., & Sangster, C. (2007). Development of metacognition and self-regulated learning in young children: Role of collaborative and peer-assisted learning. *Journal of Cognitive Education and Psychology*, *6*(3), 433–455. https://doi.org/10.1891/194589507787382043

Wisniewski, E. J. (1995). Prior knowledge and functionally relevant features in concept learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(2), 449–468. https://doi.org/10.1037/0278-7393.21.2.449