Building Abstraction: The Role of Representation and Structural Alignment in Learning Causal System Categories

Margarita Pavlova (mvpavlova.nbu.bg)
Department of Cognitive Science and Psychology, New Bulgarian University, 21 Montevideo Street
Sofia 1618, Bulgaria

Abstract

The present study examined the role of detecting the initial causal system model followed by engaging in active vs. passive structural alignment in recognizing the key causal principles in subsequent novel examples. The results echo prior research on the benefit of analogical comparison in learning relational categories: participants who were prompted to compare outperformed participants in the baseline condition. Moreover, while the accurate representation of the causal system predicted noticing the relational structure in novel examples, making more accurate relational mappings made participants more likely to notice the structure above and beyond having an accurate representation. These findings offer insight into the role of active vs. passive analogical comparison and have implications for conditions that might support learning of relational categories.

Keywords: relational reasoning; learning; transfer; relational categories

Introduction

A key goal of education is building transferable knowledge that learners can apply in different situations and beyond the specific learning context in which they obtained it (National Research Council, 2012; OECD, 2016). Such learning requires appreciating the underlying common relational patters among disparate situations. Moreover, attending to relational information is critical to cross-domain transfer (Gentner, 1983). Novice learners struggle to recognize these patterns because their representations often do not incorporate the relational-structural information in a way that is independent of the various learning situations which contributes to their inert knowledge (Fries et al., 2021). Subsequently, they tend to base similarity judgements and categorization on surface similarities (e.g., Chi et al., 1981; Gentner et al., 1993) and struggle to generalize the material to novel situations particularly when there are different surface characteristics that might lure them toward incorrect inferences (e.g., Gick & Holyoak, 1980, 1983; Holyoak & Koh, 1987; Novick, 1988; Trench & Minervino, 2015). In that sense, supporting the building of abstract relational representations that connect disparate situations and domains seems a fruitful avenue in providing education and training that builds transferable knowledge (Goldwater & Schalk, 2016). One way to achieve abstract relational representations is via structural alignment, the process by which the relational structures of two situations are aligned and compared and inferences are carried out from the better known one to the lesser known one.

Teaching Relational Reasoning

Noticing and applying the underlying relational structure among various relational categories is a sign of expertise. Structural alignment supports the acquisition of expert-like domain knowledge and the construction of schemas (Alfieri et al., 2013; Gick & Holyoak, 1980, 1983) and supports the organization and representation of knowledge in more coherent relational systems which learners can flexibly manipulate according to the situation. Structural alignment also helps to change the mental representations of one or both situations. Thus, relational reasoning can be seen as a means as well as an outcome of learning (Richland & Simms, 2015).

However, learning by analogy is challenging and can be sensitive to the conditions of the learning context. Representing instructional examples as systems of relations, aligning, mapping, and drawing inferences based on these systems draws on executive functions such as working memory and inhibitory control (Krawczyk et al., 2008; Richland & Simms, 2015); relational reasoning correlates with individual differences in fluid intelligence (Gray & Holyoak, 2020; Kubricht et al., 2017; Vendetti et al., 2014). Relatedly, transfer is difficult. People often fail to notice or retrieve the appropriate relational information because the current situation shares few object or domain similarities with the previously encountered information (Gentner et al., 1993; Gick & Holyoak, 1980, 1983; Holyoak & Koh, 1987).

On the other hand, recent research suggests that under appropriate learning context, relational reasoning can be trained. For example, Kubricht et al. (2017) showed that presenting learners with instructional materials that facilitate comprehension, mediates analogical transfer. Goldwater & Gentner (2015) found that interventions that facilitate categorization based on relational structure can lead to recognizing that schema in novel situations. Most recently, Kessler et al. (2023) showed that training by capitalizing on the construction of relational category knowledge can lead to an ability to recognize the key causal structure and subsequently relates to higher performance on complex problem-solving. Taken together, these results suggest that relational reasoning seems to not be entirely dependent on inter-individual differences: it can be practiced and trained.

Acquiring Expert-like Knowledge

A key difference between experts and novices seems to be in the way they organize their knowledge. Experts tend to organize their knowledge in terms of relational-structural similarity. For example, in problem/example sorting tasks,

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

experts tend to group examples into relational categories such as physics "energy" problems or "positive feedback" examples, whereas novices tend to group them into feature-based categories such as physics "pulley" problems or "biodiversity" examples (Chi et al., 1981; Rottman et al., 2012).

Prior knowledge interacts with analogical reasoning and the ability to extract relevant relations. Experts are thought to have a propensity to notice relational similarities between disparate domains (Dunbar, 2001; Goldwater et al., 2021; Novick, 1988) and to habitually encode new examples in the terms of key relational principles in the domain which contributes to their ability to retrieve relational matches in subsequent transfer tasks (Goldwater et al., 2021).

Learning Relational Categories

Members of relational categories share a common relational structure (Markman & Stilwell, 2001). Relational categories can be divided into role-governed categories which specify that members play the same role in a global relational structure and schema-governed categories which specify relational systems (Markman & Stilwell, 2001 and see Gentner & Kurtz, 2005 for a similar discussion). Importantly, categorizing phenomena based on an underlying relational principle can activate conceptual knowledge about how these phenomena exemplify the principle and enable inferences of how that principal works in subsequent transfer. Conversely, categorizing phenomena based on surface features is unlikely to activate such conceptual knowledge and inferences and instead can lead to negative transfer (Novick, 1988). In sum, categorizing by relational principles activates schemas which can be useful in transfer when novel situations share relational-structural but not surface information (Novick, 1988).

Using an ambiguous card sorting task (ACST), Rottman and colleagues (2012) demonstrated that experts but not novices spontaneously group novel phenomena in terms of a shared relational schema. The researchers concluded that due to their broader cross-domain knowledge and potential opportunities to abstract an underlying general principle from multiple examples, experts tend to notice and predict key relational phenomena.

A related line of research has investigated whether instructional and training conditions can bring about expert-like knowledge. For example, Goldwater & Gentner (2015) found that providing maximum instructional support (e.g., full explication of how learning examples fit a causal system) coupled with structural alignment lead to more causal sorts in the ACST. Importantly, there was an "added" benefit of structural alignment: even learners who had accurate representations of the learning examples benefited from analogical comparison compared to those who were not prompted to compare learning examples and more reliably recognized the key relational schemas in novel examples.

Building on that work, Pavlova and Greenhoot (2023) found that more minimal instructional support such as relational labels and short definitions coupled with analogical comparison support recognizing the relational schemas in novel examples. Their results echo research on relational labels that providing labels of the relational schema can support learning and transfer because they invite comparison of the learning examples and boost transfer (Goldwater & Jamrozik, 2019; Jamrozik & Gentner, 2020). In addition, relational labels can support the re-representation of the material in terms of a more general relational schema by promoting a uniform relational encoding (Gentner, 2010).

Recently, Kessler et al. (2023) showed that instructional and training materials which support building abstract schemas of relational categories supports acquiring expert-like knowledge. In their study, participants first sorted example phenomena using the ACST (Rottman et al., 2012) and then underwent different types of training. The training consisted of intervention of providing full explication and prompting structural alignment (following Goldwater & Gentner, 2015) and a tutorial capitalizing on procedural knowledge and conceptual understanding of the causal models. The results showed that participants who responded to the training and successfully shifted to more relational sorting (as measured by a parallel ASCT at post-test), also performed better on a subsequent complex problem-solving task. These results suggest that training of categorization can promote transfer, particularly when the learners have understood the key concept in a given domain. The researchers concluded that categorization training can promote transfer because it allows learners to organize their knowledge in a way that makes the relevant knowledge more accessible (Kessler et al., 2023).

Overall, the results from these studies are related to a recent view on the benefit of categorization for spontaneous transfer proposed by Kurtz & Honke (2020) – the *category status hypothesis* according to which "to the extent that the form of a knowledge representation is more category-like, the knowledge will be easier to access under the critical conditions of high structural match and low superficial match" (Kurtz & Honke, 2020, p. 805). The researchers found that learners who were prompted to categorize examples based on a shared relational principle outperformed learners who were prompted to compare pairs of examples and extract the relational principle (e.g., abstraction by comparison) on a subsequent transfer task. Kurtz & Honke (2020) interpreted these results in accordance with the category status hypothesis that having encoded the learning material as a relational category, aids learning and transfer by activating the relevant conceptual knowledge, making the representations more abstract, but importantly, also providing a retrieval path for finding appropriate knowledge later.

Relational categories bridge disparate domains and are thus crucial for learning and in educational contexts (Goldwater & Schalk, 2016). Learning relational categories is critical for relational discovery and far transfer, which is a fundamental goal and outcome of education (Don et al., 2023). A growing body of research looks at the role of relational categorization and continues to explore the favorable conditions of supporting relational categorization and building expert-like knowledge in learners.

Active vs. Passive Learning

A large body of research illustrating the benefits of active learning (e.g., Freeman et al., 2014; Theobald et al., 2020) suggests that when instructors offer active learning opportunities, students show greater gains in learning and are at lower risk for failing. Conversely, using exposition-centered methods (i.e., lecturing) is less effective than constructivist approaches in which students engage in working with the material.

In their review, Gureckis & Markant (2012) propose that active learning is better than passive learning because it allows learners to choose which information to receive at instruction, thus restricting the search space for hypotheses or set sampling and leads to more efficient strategies. On the other hand, active learning is related to the learner's quality of representation of the material: if the learner's representation is flawed, that might lead to biased and ineffective learning (Gureckis & Markant, 2012). Therefore, providing a source of accurate representations (e.g., in the form of instructional material, or expert model) might support learning while still capitalizing on active exploration. In a related study, MacDonald & Frank (2016; see also Markant & Gureckis, 2014) showed that while opportunities for active learning boost the effectiveness of learning, providing passive learning prior to active learning actually leads to more advantageous learning. The researchers concluded that passive learning might act to constrain the search space for strategies and thus allows learners to explore more efficiently. Relatedly, work on *productive failure* (Kapur, 2009; Westermann & Rummel, 2012) shows that providing an opportunity for students to struggle with a task can aid subsequent instruction and learning. In doing so, learners might engage in compensatory strategies to fill gaps in their knowledge which in turn might prepare them for future learning.

A review by Chi (2009) provides a framework for active, constructive, and interactive learning and suggests that engaging in active and constructive vs. passive activities promotes better learning. Active activities include processes such as activating prior knowledge, encoding, and storing new information, and searching existing knowledge space. Constructive activities include processes such as inferring new knowledge, integrating novel information with existing information, organizing one's knowledge to make it more coherent, repairing one's false beliefs, restructuring one's knowledge (Chi, 2009). According to this framework, drawing analogies, comparing and contrasting cases, drawing conceptual models are constructive activities.

Generating information leads to better learning compared to passive learning. This effect is called the *generation effect* (Bertsch et al., 2007; Metcalfe & Kornell, 2007) and is associated with better recall and greater likelihood that learners will notice relevant relations, for example by trying explain how exemplars belong to a given category (e.g., Edwards et al., 2019). Generating solutions to analogies unlocks a relational mindset and predicts performance on relational mapping tasks (Vendetti et al., 2014) and retrieval of analogous examples (Goldwater & Jamrozik, 2019).

Relating this to the analogical learning research, generally, structural alignment is conceptualized as an active process, and not merely a juxtaposition of analogs (Catrambone & Holyoak, 1989; Gentner et al., 2003; Kurtz et al., 2001). While some work demonstrates that engaging in more elaborate analogical comparison (e.g., Goldwater & Gentner, 2015; Kessler et al., 2023; Kurtz et al., 2001) boosts transfer, the role of active vs. passive structural alignment has not been systematically studied. This study is an attempt to launch a research line exploring the specific conditions that make active and passive learning beneficial to learning and transfer.

The present study sought to investigate whether training with active vs. passive structural alignment would (1) boost subsequent transfer in a relational category sorting task and (2) whether the effect of structural alignment would be in addition to having understood the conceptual model of the causal systems (thus replicating findings by Goldwater & Gentner, 2015). It was hypothesized that active comparison, as engaging in an active and constructive type of learning (Chi, 2009; Fonseca & Chi, 2010), would support transfer of the relational-causal information above and beyond the gains of gleaning the causal-relational structure from reading full explications of the learning examples.

Method

Participants

One hundred and seven NBU students (M_{age} = 29.64; *SD* = 11.08; 61 females) took part in the study. They were assigned to one of three conditions as follows: 37 in Active Comparison, 34 in Passive Comparison, and 36 in Baseline. Additionally, data from 31 students were removed from analyses due to them failing at least one of the two attention checks. The data were collected online via Google Forms between Spring and Fall 2023. Students received partial course credit for participation.

Design and Procedure

A factorial design with 3 conditions (Active Comparison, Passive Comparison, and Baseline condition) was used; all variables were manipulated between participants. All participants except those in the Baseline condition were first presented with 10 short examples depicting 5 causal systems (common cause, common effect positive feedback, negative feedback, and causal chain). Each example was followed by a label, definition, and full explication. After reading each, participants had to choose a diagram which best depicted the causal system in the example. Then, participants in the comparison groups were presented with the same examples in pairs and were either provided a mapping table with corresponding elements from each example (Passive Comparison) or were asked to fill a mapping table (Active Comparison) where the elements from the 2 examples were provided and participants had to indicate the correct mapping between them (i.e., they had to indicate which element of Situation 1 corresponded to which element of Situation 2). Next, all participants proceeded to the Transfer Phase where they solved an ambiguous card sorting task. Participants in the Baseline condition skipped the Learning Phase and completed only the Transfer Phase.

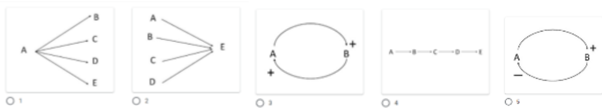

Figure 1: Causal diagrams for each of the five causal systems.

a)

Situation 1	Situation 2
quality of cell phone reception	1
distance from a cell tower	2
terrain (mountains and buildings)	3
other electric devices	4
weather	5

	biodiversity	availability of resources (food and habitat)	climate change	poaching and hunting	invasive species
quality of cell phone reception	○	○	○	○	○
distance from a cell tower	○	○	○	○	○
terrain (mountains and buildings)	○	○	○	○	○
other electric devices	○	○	○	○	○
weather	○	○	○	○	○

b)

Situation 1	Situation 2
quality of cell phone reception	biodiversity
distance from a cell tower	availability of resources (food and habitat)
terrain (mountains and buildings)	climate change
other electric devices	poaching and hunting
weather	invasive species

○ Understood

Figure 2: Example of the mapping table. Panel a) Mapping table for the Active Comparison condition; panel b) Mapping table for the Passive Comparison condition.

Materials

Instruction Materials The instruction materials consisted of 10 short examples depicting 5 causal systems from two domains (e.g., electrical engineering and biodiversity; adapted from Rottman et al., 2012). For the first part of the Learning Phase, where participants had to select a diagram, 5 diagrams depicting the 5 causal systems were drawn (Figure 1). For the second part of the Learning Phase, a mapping table was designed (adapted from Kurtz et al., 2001). Participants in the Active Comparison condition had to indicate the corresponding elements between Situation 1 and Situation 2 (Figure 2, panel a). Participants in the Passive Comparison condition had to read the mapping table and indicate they have done so by selecting an "Understood" button below it (Figure 2, panel b).

Assessment Materials Assessment materials consisted of the Ambiguous Card Sorting Task (ACST; adapted from Rottman et al., 2012). The materials were translated into Bulgarian using independent backwards and forwards translations. The ACST consisted of 20 example phenomena composing a matrix of five causal systems crossed with five content domains. Because there were two types of sorts (e.g., according to domain or causal system), participants sorted the examples twice.

Results

The data were cleaned and processed using R version 4.3.0 (R Core Team, 2023) and statistical analyses were conducted in JASP version 0.16.3.0 (JASP Team, 2022). For each correctly identified diagram, participants received 1 point; thus, diagram accuracy was calculated as a proportion of correctly identified diagrams. Similarly, for each correctly selected element in the mapping task, participants (in Active Comparison) received 1 point; thus, relational mapping accuracy was calculated as a proportion of correctly selected relational matches. For each sorting, the types of sorting (causal, domain, error) were calculated as a proportion of the total number of cards sorted.

Effects of Structural Alignment

The first set of analyses were conducted among the 3 groups without considering the rate of accurately recognized diagrams and the accuracy of the mapping task. The aim was to see how the comparison groups perform relative to the baseline group.

A repeated measures ANOVA on the rate of **causal sorts** between Sorting (Sort 1 and Sort 2) as within-subjects factor and Group as between-subjects factor was conducted. The Levene's test for equal variances showed a violation of equality of variances in Sorting ($p < .05$). Nevertheless, post-hoc tests with Bonferroni corrections were conducted since these tests are independent from this assumption (Hsu, 1996). These revealed significantly more causal sorts in the comparison groups compared to the baseline group: Active Comparison vs. Baseline: $t(106) = 3.769$, $p < .001$; Passive Comparison vs. Baseline: $t(106) = 3.120$, $p = .007$. There was no difference between the two comparison groups. A Friedman Test revealed no difference between Sort 1 and Sort 2 ($p = .210$).

A repeated measures ANOVA on the rate of **domain sorts** between Sorting (Sort 1 and Sort 2) as within-subjects factor and Group as between-subjects factor revealed a main effect of Sorting: $F(1, 104) = 4.509$, $p = .036$, $\eta^2 = .005$ and a main effect of Group: $F(2, 104) = 13.639$, $p < .001$, $\eta^2 = .183$. There was no interaction between Sorting and Group ($F(2, 104) = .599$, $p = .551$. Post-hoc tests with Bonferroni correction revealed that there were more domain sorts in the second sort: $t(106) = -2.124$, $p = .036$. These also revealed significantly more domain sorts in the baseline group compared with the comparison groups: Active Comparison vs. Baseline: $t(106) = -4.808$, $p < .001$; Passive Comparison vs. Baseline: $t(106) = -4.156$, $p < .001$.

A repeated measures ANOVA on the rate of **error sorts** between Sorting (Sort 1 and Sort 2) as within-subjects factor and Group as between-subjects factor revealed no main effects, nor interaction: Sorting: $F(1,104) = 2.032$, $p = .157$; Group: $F(2,104) = .210$, $p = .811$; interaction: $F(2,104) = 1.289$; $p = .280$.

Effects of the Causal System Representation in Addition to Structural Alignment

The next set of analyses are linear regressions and consider only the first sorting. These analyses were conducted among the two comparison groups only, since the participants in these groups completed the Learning Phase. The goal of these analyses was to understand whether the accuracy of the representation of causal systems of the learning examples predicts recognizing the causal systems in novel examples.

A linear regression predicting **causal sorts** from diagram accuracy and group (active or passive comparison) revealed that the proportion of accurately selected diagrams predicted *more* causal sorts: $F(1,68) = 6.883$, $p = .002$, $R^2 = .168$; $t(68) = 3.709$, $p < .001$. The group was not a significant predictor: $t(68) = .454$, $p = .651$. A separate linear regression on the **domain sorts** from diagram accuracy and group (active or passive comparison) revealed that the proportion of accurately selected diagrams predicted *less* domain sorts: $F(1,68) = 5.485$, $p = .006$, $R^2 = .139$; $t(68) = -3.244$, $p = .002$. The group was not a significant predictor: $t(68) = .169$, $p = .866$. Finally, a linear regression on the rate of **error sorts** revealed that the model was not significant: $F(2,68) = 1.534$, $p = .223$.

These results suggest that regardless of being prompted to compare examples actively or passively, the reliable predictor in recognizing the causal system in novel examples (i.e., sorting more causally or less by domain) is the accuracy of the initial causal system representation of the learning examples (see Table 2 for the types of sorts among the groups).

Effects of Relational Mapping in Addition to Causal System Representation

The final set of analyses were conducted withing the Active Comparison group and examined whether the established effect of accuracy of the representation of the causal system initially boosted performance in the mapping task, and whether there were combined effects of representation and mapping on subsequent causal sorting.

To better understand the role of the causal model representation and structural alignment in recognizing the causal model in novel situations, two separate hierarchical linear regression models on the rate of causal and domain sorts in the first sorting with Diagram Accuracy and Relational Mapping Accuracy as predictors were conducted. Model 1 included Diagram Accuracy only and Model 2 included the Relational Mapping Accuracy added as a predictor together with Diagram Accuracy.

Regressing **causal sorts** on diagram accuracy showed that Model 1 was significant ($F(1, 35) = 8.341$, $p = .007$) and explained 19.2% of the variance in the causal sorts. Regressing causal sorts on diagram accuracy and mapping accuracy showed that Model 2 was also significant ($F(2, 34) = 7.801$, $p = .002$) and explained 31.5% of the variance in the causal sorts and registered a significant change in R^2 ($\Delta R^2 = .122$, $p = .019$). Regressing **domain sorts** on diagram accuracy showed that Model 1 was significant ($F(1, 35) = 6.049$, $p = .019$) and explained 14.7% of the variance in the domain sorts. Regressing domain sorts on diagram accuracy and mapping accuracy showed that Model 2 was also significant ($F(2, 34) = 8.027$, $p = .001$) and explained 28.1% of the variance in the domain sorts and registered a significant change in R^2 ($\Delta R^2 = .173$, $p = .006$). See Table 3 for the regression analyses.

These results suggest that accurately matching relational correspondences across the learning examples predicts subsequent recognition of the key causal structure above and beyond the accurate detection of the causal schema from the learning examples.

Table 2: Means and Standard Deviations of Types of Sorts among the Groups

Group	Type of Sort	Sort 1 *M*	Sort 1 *SD*	Sort 2 *M*	Sort 2 *SD*
Active Comparison	Causal	0.319	0.216	0.335	0.233
	Domain	0.253	0.178	0.273	0.201
	Error	0.428	0.175	0.392	0.171
Passive Comparison	Causal	0.313	0.257	0.284	0.247
	Domain	0.278	0.159	0.296	0.195
	Error	0.409	0.176	0.421	0.187
Baseline	Causal	0.151	0.177	0.133	0.17
	Domain	0.442	0.202	0.493	0.23
	Error	0.407	0.18	0.374	0.207

Discussion

The present study examined the role of comprehension of causal systems models and structural alignment in subsequently recognizing key relational principles in novel examples. The results showed that training in structural alignment (passive or active) leads to more causal sorts compared to baseline. Across the full dataset, this advantage was explained by the accuracy of the initial causal model representation (e.g., diagram accuracy). Importantly, within the active comparison group, participants who had made accurate relational mappings in addition to having accurate causal model representations, were more likely to recognize the causal system in novel examples. Taken together, these results support and extend prior work on the benefit of analogical comparison on learning and transfer (Alfieri et al., 2013; Gadgil et al., 2012; Gick & Holyoak, 1980, 1983; Goldwater & Gentner, 2015; Kessler et al., 2023; Kurtz et al., 2001; Loewenstein et al., 2003).

The current findings also accord with *the category status hypothesis* (Kurtz & Honke, 2020). The present study found that accuracy of the causal systems diagrams predicted causal sorts in the ACST. Identifying the correct diagram can serve as evidence of having a more accurate representation of the relational structure in the learning examples. Thus, it is possible that these participants might have learned the learning examples as “types of relational categories” which allowed them to encode the phenomena described as examples of such categories and subsequently recognize it in novel situations.

A somewhat surprising finding was that, overall, it seemed that there was no difference between the active and passive comparison. One explanation might be that conducting an online unmoderated study might have attenuated the positive effects of comparison. It is possible that participants were less engaged with the learning material. Alternatively, it is possible that the processes of having to read vs. select matching elements in the mapping task are not that different. More work is needed to systematically examine the role of exposure vs. engagement in structural alignment in learning and transfer. Nonetheless, the results from within the active comparison group support the notion that actively engaging in the material boosts learning and transfer. Indeed, participants who responded to the training in the intended way and selected more correct corresponding relational elements, were more likely to recognize the key relational patterns in novel situations above and beyond the accurate detection of the causal system in the learning examples.

A few limitations of this study are worth noting. First, the learning phase omitted a crucial aspect of generating the underlying principle between the two learning examples (e.g., as in Goldwater & Gentner, 2015). In future work we plan to include a multistep abstraction by comparison procedure (e.g., Kurtz & Honke, 2020) during learning to boost the conceptual understanding of the key phenomena and potentially boost transfer to novel situations. Second, when participants selected the diagrams, the labels, together with the explications were visible. It is thus possible that they simply matched the diagram to the provided information and did not have to rely on extracting the structure from the learning examples themselves. In future work, we plan to test the representation of the causal models directly following exposure to the learning examples prior to any additional information. This would provide a more accurate test of the learners’ initial causal model representations.

Table 3: Regression Coefficients of Diagram and Relational Mapping Accuracy on Causal and Domain Sorts

A: Causal Sorts

Variable	Model 1			Model 2		
	B	β	*SE*	*B*	β	*SE*
Constant	-0.13		0.159	-0.225		0.153
Diagram	0.511^{**}	0.439	0.177	0.397	0.341^{**}	0.172
Relational Mapping				0.405	0.363^{**}	0.165
R^2	0.192				0.315	
ΔR^2					0.122	

B: Domain Sorts

Variable	Model 1			Model 2		
	B	β	*SE*	*B*	β	*SE*
Constant	0.576^{***}		0.134	0.67^{***}		0.126
Diagram	-0.368^{**}	-0.384	0.15	-0.256	-0.267	0.141
Relational Mapping				-0.398	-0.432^{**}	0.135
R^2	0.147				0.321	
ΔR^2					0.173	

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. Panel A: Causal Sorts in Sort 1. Panel B: Domain Sorts in Sort 1. Model 1 includes only Diagram Accuracy as predictor; in Model 2, Mapping Accuracy was added as a predictor.

Acknowledgements

I thank the participants in the study and the research assistants Veronika Penkova, Kattrin Katovska, Steliana Yankova, Pamela Pavlova, Velina Ivanova, Yoana Vutova, Bogdana Stoycheva, and Loreta Stoyanova for their help in data collection. I am grateful to Victoria Kuneva for the initial translation and to Velina and Veronika for the backwards translation of the materials. I also thank Penka Hristova for her feedback and discussions of the study design. Finally, I thank Garrett Honke and two anonymous reviewers for their valuable feedback and comments on an initial draft of this paper.

References

Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning Through Case Comparisons: A Meta-Analytic Review. *Educational Psychologist*, *48*(2), 87–113.

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition*, *35*(2), 201–210.

Catrambone, R., & Holyoak, K. J. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(6), 1147–1156.

Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, *1*(1), 73–105.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*, 121–152.

Don, H. J., Goldwater, M. B., & Livesey, E. J. (2023). Cognition of relational discovery: Why it matters for effective far transfer and effective education? *Frontiers in Psychology*, *14*, 957517.

Dunbar, K. (2001). The analogical paradox: Why analogy is so easy in naturalistic settings yet so difficult in the psychological laboratory. In D. Gentner, K. J. Holyoak, & B. Kokinov (Eds.), *The analogical mind: Perspectives from cognitive science*. The MIT Press.

Edwards, B. J., Williams, J., Gentner, D., & Lombrozo, T. (2019). Explanation recruits comparison in a category-learning task. *Cognition*, *185*, 21–38.

Fonseca, B. A., & Chi, M. T. H. (2010). Instruction based on self-explanation. In *The handbook of research on learning and instruction*. Routledge.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, *111*(23), 8410–8415.

Fries, L., Son, J. Y., Givvin, K. B., & Stigler, J. W. (2021). Practicing Connections: A Framework to Guide Instructional Design for Developing Understanding in Complex Domains. *Educational Psychology Review*, *33*(2), 739–762.

Gadgil, S., Nokes-Malach, T. J., & Chi, M. T. H. (2012). Effectiveness of holistic mental model confrontation in driving conceptual change. *Learning and Instruction*, *22*(1), 47–61.

Gentner, D. (1983). Structure-Mapping: A Theoretical Framework for Analogy. *Cognitive Science*, *7*(2), 155–170.

Gentner, D. (2010). Bootstrapping the Mind: Analogical Processes and Symbol Systems. *Cognitive Science*, *34*(5), 752–775.

Gentner, D., & Kurtz, K. J. (2005). Learning and using relational categories. In W. K. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman, & P. W. Wolff (Eds.), *Categorization inside and outside the laboratory*. American Psychological Association.

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology*, *95*(2), 393–408.

Gentner, D., Rattermann, M. J., & Forbus, K., D. (1993). The roles of similarity in transfer: Separating retrievability from inferential soundness. *Cognitive Psychology*, *25*, 524–575.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, *12*, 306–355.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*(1), 1–38.

Goldwater, M. B., & Gentner, D. (2015). On the acquisition of abstract knowledge: Structural alignment and explication in learning causal system categories. *Cognition*, *137*, 137–153.

Goldwater, M. B., Gentner, D., LaDue, N. D., & Libarkin, J. C. (2021). Analogy Generation in Science Experts and Novices. *Cognitive Science*, *45*(9).

Goldwater, M. B., & Jamrozik, A. (2019). Can a relational mindset boost analogical retrieval? *Cognitive Research: Principles and Implications*, *4*(1), 47.

Goldwater, M. B., & Schalk, L. (2016). Relational categories as a bridge between cognitive and educational research. *Psychological Bulletin*, *142 7*, 729–757.

Gray, M. E., & Holyoak, K. J. (2020). Individual differences in relational reasoning. *Memory & Cognition*, *48*(1), 96–110.

Gureckis, T. M., & Markant, D. B. (2012). Self-Directed Learning: A Cognitive and Computational Perspective. *Perspectives on Psychological Science*, *7*(5), 464–481.

Holyoak, K. J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory & Cognition*, *15*(4), 332–340.

Hsu, J. (1996). *Multiple Comparisons: Theory and Methods* (1 ed.). Chapman and Hall/CRC.

Jamrozik, A., & Gentner, D. (2020). Relational labeling unlocks inert knowledge. *Cognition*, *196*, 104–146.

JASP Team. (2022). *JASP* (0.16.3) [Computer software]. University of Amsterdam. https://jasp-stats.org/

Kapur, M. (2009). Productive failure. *Cognition and Instruction*, *26*(3), 379–425.

Kessler, F., Proske, A., Urbas, L., Goldwater, M., Krieger, F., Greiff, S., & Narciss, S. (2023). Promoting Complex Problem Solving by Introducing Schema-Governed Categories of Key Causal Models. *Behavioral Sciences*, *13*(9), 701.

Krawczyk, D. C., Morrison, R. G., Viskontas, I., Holyoak, K. J., Chow, T. W., Mendez, M. F., Miller, B. L., & Knowlton, B. J. (2008). Distraction during relational reasoning: The role of prefrontal cortex in interference control. *Neuropsychologia*, *46*(7), 2020–2032.

Kubricht, J. R., Lu, H., & Holyoak, K. J. (2017). Individual differences in spontaneous analogical transfer. *Memory & Cognition*, *45*(4), 576–588.

Kurtz, K. J., & Honke, G. (2020). Sorting out the problem of inert knowledge: Category construction to promote spontaneous transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(5), 803–821.

Kurtz, K. J., Miao, C.-H., & Gentner, D. (2001). Learning by analogical bootstrapping. *The Journal of the Learning Sciences*, *10*(4), 417–446.

Loewenstein, J., Thompson, L., & Gentner, D. (2003). Analogical Learning in Negotiation Teams: Comparing Cases Promotes Learning and Transfer. *Academy of Management Learning and Education*, *2*(2), 119–127.

MacDonald, K., & Frank, M. C. (2016). When does passive learning improve the effectiveness of active learning? *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*.

Markant, D. B., & Gureckis, T. M. (2014). Is it better to select or to receive? Learning via active and passive hypothesis testing. *Journal of Experimental Psychology: General*, *143*(1), 94–122.

Markman, A. B., & Stilwell, C. H. (2001). Role-governed categories. *Journal of Experimental & Theoretical Artificial Intelligence*, *13*(4), 329–358.

Metcalfe, J., & Kornell, N. (2007). Principles of cognitive science in education: The effects of generation, errors, and feedback. *Psychonomic Bulletin & Review*, *14*(2), 225–229.

National Research Council. (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st century*. The National Academies Press.

Novick, L. R. (1988). Analogical transfer, problem similarity, and expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 510–520.

OECD. (2016). *PISA 2015 Results (Volume I): Excellence and Equity in Education*. OECD.

Pavlova, M., & Greenhoot, A. F. (2023). Comparison and Explanation in Learning Causal System Categories. In M. B. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society* (pp. 1499–1507). Cognitive Science Society.

R Core Team. (2023). *R: A Language and Environment for Statistical Computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Richland, L. E., & Simms, N. (2015). Analogy, higher order thinking, and education: Analogy, higher order thinking, and education. *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*(2), 177–192.

Rottman, B. M., Gentner, D., & Goldwater, M. B. (2012). Causal Systems Categories: Differences in Novice and Expert Categorization of Causal Phenomena. *Cognitive Science*, *36*(5), 919–932.

Theobald, E. J., Hill, M. J., Tran, E., Agrawal, S., Arroyo, E. N., Behling, S., Chambwe, N., Cintrón, D. L., Cooper, J. D., Dunster, G., Grummer, J. A., Hennessey, K., Hsiao, J., Iranon, N., Jones, L., Jordt, H., Keller, M., Lacey, M. E., Littlefield, C. E., … Freeman, S. (2020). Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. *Proceedings of the National Academy of Sciences*, *117*(12), 6476–6483.

Trench, M., & Minervino, R. A. (2015). The Role of Surface Similarity in Analogical Retrieval: Bridging the Gap Between the Naturalistic and the Experimental Traditions. *Cognitive Science*, *39*(6), 1292–1319.

Vendetti, M. S., Wu, A., & Holyoak, K. J. (2014). Far-Out Thinking: Generating Solutions to Distant Analogies Promotes Relational Thinking. *Psychological Science*, *25*(4), 928–933.

Westermann, K., & Rummel, N. (2012). Delaying instruction: Evidence from a study in a university relearning setting. *Instructional Science*, *40*(4), 673–689.