## Title
Defending Diversity: Providing Examples from Different Domains Enhances Application of System Principles Beyond the Domains Covered by the Examples



**Journal**
Proceedings of the Annual Meeting of the Cognitive Science Society, 44(44)

**Authors**
D'Angelo, Verónica S
Trench, Maximo

**Publication Date**
2022

Peer reviewed

# Defending Diversity: Providing Examples from Different Domains Enhances Application of System Principles Beyond the Domains Covered by the Examples

**Verónica D'Angelo (VERODANGELO71@Gmail.Com)**
Instituto Rosario de Investigaciones en Ciencias de la Educación
IRICE-CONICET/UNR

**Máximo Trench (MAXIMO.TRENCH@Crub.Uncoma.Edu.Ar)**
Department of Psychology, Universidad Nacional del Comahue
IPEHCS-CONICET/UNCo

## Abstract

The external provision of examples has proven the most successful approach to aid learning and application of declarative concepts (i.e., abstract concepts denoted by key terms and short definitions that can be applied to a wide variety of scenarios). The current experiment sought to further this line of research by exploring the effect of using thematically varied examples on learners' ability to classify novel exemplars and near misses of five *system principles* that cut across thematic domains. Results revealed that thematic variation increased learners' ability to reject near-misses and, more crucially, to classify novel exemplars from domains not covered by the studied examples. The fact that this enhanced flexibility was unaccompanied by poorer performance in rejecting near misses or classifying new items from domains covered by the learned examples renders this strategy readily applicable in instructional settings. We end by discussing possible mechanisms that could potentially explain the observed advantage of thematically varied examples.

**Keywords:** thematic variation; examples; system principles; analogy; learning.

## Introduction

A major goal of education is that students can apply their knowledge to contents (and in contexts) different from those of the initial learning (Barnett & Ceci, 2002; Day & Goldstone, 2012). Among the different kinds of contents that can be potentially transferred (e.g., reasoning strategies, complex multistep procedures, motor skills, etc.), declarative concepts are prevalent across a variety of educational domains and degrees of expertise. They consist in abstract concepts denoted by key terms and short definitions—usually one or two sentences long—that can be applied to a wide variety of scenarios (Rawson, Thomas, & Jacoby, 2015). As opposed to entity categories like "table", which are commonly represented by independent features, declarative concepts often consist in relational categories (Markman & Stilwell, 2001), which tend to be defined by a set of interrelations rather than static features.

During declarative concept learning, concept definitions are often treated as the primary targets of learning, under the assumption that once the learner has assimilated the abstract relationships that hold among the constituent elements of the concept, students can rely on the definition for prototypical concept application activities such as classifying novel exemplars (Zamary, 2019). But abstract definitions can sometimes be difficult for students to understand. To facilitate the acquisition of declarative concepts, textbooks and instructors frequently provide students with concrete examples. Presenting examples leads to better application of declarative concepts than allotting the same amount of time to restudy the definitions (Balch, 2005; Rawson et al., 2015).

The particular mechanisms through which examples facilitate the acquisition and application of declarative concepts are still unclear. One possibility is that they help clarify the schemas conveyed by the definitions. However, an alternative possibility is that examples may directly support concept application, to the extent that they can be used as base analogs against which the new exemplars are matched through a process of analogical mapping (Ross, 1987). These alternatives need not be mutually exclusive, since knowledge about the examples and knowledge about the definition might serve as a fallback option when the other kind fails (the *dual-route hypothesis*, Zamary, 2019).

If examples are so important for declarative concept learning, how to better employ them during instruction? Counter to constructivists' expectations, the simple presentation of a small set of examples (e.g., 3 or 4) has proven superior than learning schedules requiring a more active role on the part of the learner, such asking participants to generate their own examples (Zamary & Rawson, 2018a) or to complete faded examples (Zamary & Rawson, 2018b). Despite the superiority of the provision of examples compared to competing strategies, the learning outcomes obtained through this strategy are still far from ceiling (see also Zhang, 2019). This situation becomes even more problematic regarding the learning of *system principles* such as "negative feedback" or "refractory period" (Goldstone & Wilensky, 2008; Day, Goldstone & Hills, 2010), a particular kind of declarative concepts that are especially relevant for science learning, as they cut across thematic domains. Recent evidence suggests that explicating how each of the learning examples instantiates the schema conveyed by the definition can boost classification of further examples from domains not covered by the examples, but only when combined with an explicit comparison between the learning examples (Goldwater & Gentner, 2015). Hence, a pressing research question concerns how to select examples in a way that maximizes accuracy and flexibility during concept application,



In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society*. 

but under more widely applicable instructional procedures. The present study seeks to advance current research on example presentation by assessing the effects of the thematic diversity of learned examples during the application of system principles.

## Promise and Perils of Using Diverse Examples

In the field of concept learning, it has been proposed that the selection of examples with mismatching surface features may aid in highlighting the abstract features of concepts, therefore preventing the subgeneralization of their definitions (Merrill, Tennyson, & Posey, 1992). In the related field of analogical abstraction, it has likewise been suggested that as the schemas that originate in analogical comparisons tend to preserve the common elements in two situations, comparisons between overly similar examples might induce the inclusion of irrelevant intersecting features in the generated schemas, hindering application of the schema to subsequent situations with mismatching surface features (Ben-Zeev & Star, 2001; Chang, Koedinger, & Lovett, 2003). The flip side of exemplar variation might reside in an eventual difficulty for students to recognize and abstract out the relevant commonalities across examples (Gentner et al., 2003; Gick & Holyoak, 1983). In a study of example variation for learning how to solve combinatorics problems, Braithwaite and Goldstone (2015) obtained that the benefit of varied examples was less intense for students lacking prior knowledge of the concept at stake, but that it could be enhanced by enforcing attention to structural features of the problems. As we are interested in the most challenging instructional scenario wherein prior knowledge cannot be taken for granted, we accompanied each of the presented examples with an unsupervised comprehension task asking participants to match the critical elements of the example with the core components of the concept's definition.

To date, evidence supporting the provision of varied examples comes mostly from instruction on concepts and/or procedures pertaining to formal disciplines such as mathematics (Braithwaite & Goldstone; 2015, Chen & Mo, 2004; Paas & Van Merrienboer, 1994), or statistics (Chang et al., 2003; Chang, 2006; Quilici & Mayer, 1996, 2002). As discussed in more detail elsewhere (Minervino, Olguín & Trench, 2017), formal domains differ from more empirical fields of inquiry in that the properties of objects that instantiate the quantitative operations that are the target of learning tend to be rather arbitrary (e.g., dividing candy among children is hardly different from dividing cards among players), thus facilitating the suppression of irrelevant detail. In contrast, the constituent elements of empirical phenomena have properties whose relevance to the phenomenon at stake is comparatively more difficult to assess, posing challenges to their analogical extrapolation to novel cases comprising a different set of elements. Hence, the main objective of the present research consisted in assessing the advantages and disadvantages of the thematic variation of examples using a selection of abstract concepts that cut across nonformal domains as educationally relevant as physics, engineering, biology, medicine, psychology, politics, sports, or economics.

In order to assess the effects of example variation on the subsequent application of five system principles, we used a blocked schedule[1] wherein the definition of each concept was followed either by three concepts taken from one single domain, or by three examples taken from different domains. After a temporal separation of 5 to 7 days, participants received 25 target situations, which included studied items, new items from studied domains, new items from non-studied domains, near misses, and filler items. Each one was followed by a multiple-choice question listing each of the learned concepts, plus a "none of the above" option. Participants' task was to determine to which of the studied concepts each item corresponded, or else report it as not belonging to any of the learned categories.

For studied items we predicted that classification performance would be very high in absolute terms and similar across groups. This measure works mainly as a check that participants paid attention to the study examples and encoded them adequately in LTM along with their concept's definition. At the other opposite of the novelty continuum, we predicted that correct classification of novel items from non studied domains (a measure of far transfer) would be much lower overall, but comparatively higher in the varied than in the homogeneous condition. Classification of novel examples from learned domains (a measure of near transfer) was predicted to fall between the previous measures. In the event of a group difference, either numerical or significant, we anticipated that it would likely benefit the homogeneous condition, on the grounds that the presentation of three examples from a given domain would elicit a stronger association between the concept and said domain than between the concept and any of the three domains instantiated by a single example. If, on the contrary, an advantage of the varied condition on items from non-studied domains is not accompanied by a disadvantage on new examples from studied domains, there would be a clear instructional advantage of example variation. Finally, we predicted that the enhanced attention to underlying structure elicited by the varied condition would also aid participants in rejecting near-misses.

# Method

## Participants and Design

An initial sample of 60 undergraduate students volunteered to participate in the experiment and were randomly assigned to the varied examples (*N* = 28) and the homogeneous examples condition (*N* = 32). Data from three participants from the varied condition and from four participants from the homogeneous condition were discarded due to failure to complete the test phase, yielding a final sample of *N* = 25 and *N* = 28, respectively. All participants signed an informed consent for participation prior to beginning Phase 1.

[1] "Blocked" schedules not only perform better than "interleaved" schedules for this kind of concepts, but are also more representative of how examples are presented in actual instructional environments.

## Materials

Three of the five system principles selected as targets of instruction were taken or adapted from Jamrozik and Gentner (2020) [inoculation], from Goldstone and Sakamoto (2003) [competitive specialization], and from Day et al. (2010) [negative feedback]. The remaining concepts were "refractory period" and "regression to the mean". The definitions provided for each of these concepts were sufficiently abstract so as to potentially encompass instances from at least four different knowledge domains, as dictated by the experimental procedure. As an example, *refractory period* was defined as a process wherein "once certain entity has reacted to a given kind of stimulus, for a period of time it ceases to react to that same kind of stimulus". For each of these five concepts, we derived four examples from a first domain, one example from a second domain, one example from a third domain, and one example from a fourth domain (e.g., for the *refractory period* Set, we derived four examples from biology, one example from psychology, one from sports, and one from fundraising. The domains corresponding to the examples of one of the concepts did not maintain any systematic relationship with those of the other concepts, although some common domains (e.g., medicine or biology) were used in more than one concept. For each of the system principles we also derived a near-miss, which consisted in a situation that presented several structural aspects of the definition, but whose violation of a key component ultimately rendered it a non-instance of the concept. Keeping with the *refractory period* Set, the near-miss stated that "Adult humans maintain stable food preferences, and in the presence of a meal that they normally consider attractive they secrete saliva in preparation for digestion. During the first three months of pregnancy, however, hormonal processes related to gestation reduce the range of foods to which the mother is attracted, such that foods that would normally elicit a salivation response cease to generate it". While the above situation shares relevant aspects with the definition of "refractory period" (e.g., the presence of a typical reaction to particular stimuli, or a momentary suspension of said reaction to such stimuli), it violates the provided definition in that the suspension is not triggered by a previous exposure to the stimulus, but rather by an endogenous cause. Near-misses were included to assess whether the enhanced flexibility presumed to arise from thematic diversity comes at the cost of a less stringent discrimination. Finally, we included five filler items which did not maintain structural overlap with the learned concepts. Filler items were included to prevent participants from assuming that each of the test situations corresponded to one of the learned examples. Definitions were approximately 20 words long. The length of examples, near misses and fillers (range: 50-70 words) was equated across concepts, so as to prevent participants from using length as a heuristic for classifying the test items. Care was also taken in ensuring that the key words of the definitions (e.g., "stimulus", "reaction", "period", or "time" in the definition of *refractory period*) were absent from the examples, so as to preclude participants from using lexical items as a cues for classifying test items. Table 1 displays the complete definition, comprehension questions, examples, and near-miss corresponding to *negative feedback*.

## Procedure

The initial cohort of candidates were invited by the instructors of several online courses to participate anonymously in a brief online study on concept learning, which comprised two Google forms: one for the learning phase and the other for the test phase.

**Learning Phase** A first section of the form collected contact data and demographic information about age and gender. Contact information (a valid email address) was key for matching each participant's learning form (and therefore her condition) with her corresponding answers to the test form. In a subsequent section, participants watched a 3min video consisting in a screen capture where the first author demonstrated the structure of the learning trials. Instructions anticipated that participants would learn five abstract concepts, each one illustrated by three concrete examples. The video explained to participants that as comprehension of both the definition and the examples would be crucial for the testing phase, after each of the examples they would need to provide brief written answers to 3-4 questions intended to relate the central components of the definition to their corresponding real-world referent in the provided example. This was exemplified with the concept of *catalysis* (which did not belong to the study materials presented later) and each of its corresponding three examples. Each of sections 3-5 of the learning form began by presenting the definition of a concept (e.g., *negative feedback*), and followed by presenting one of the examples of said concept along with its comprehension questions. Once the last question of the third example of a concept was completed, participants proceeded to the definition of the following concept, along with their examples and their corresponding comprehension questions. The learning form thus included 17 brief sections, and took about 30 min to complete. The order of presentation of the concepts was counterbalanced by generating two versions of the learning form featuring the same concepts in different order. The only experimental manipulation concerned the selection of examples. For each of the learned concepts, while participants in the homogeneous condition received three examples all taken from the same thematic domain (e.g., three examples of *negative feedback* from the domain of engineering), participants in the varied condition received three examples from different domains (e.g., three examples of *negative feedback* taken from engineering, psychology and biology). For each concept, one of the examples of the homogeneous condition was included among the items of the varied condition.

**Application Phase** The application form was the same for both conditions. It began by informing participants that they would receive brief descriptions of situations, of which some would be instances of one of the learned concepts, and some would not. Next, it told that for each of the test situations, they would have to click on the appropriate concept, or else select "none of the above" when the situation did not belong to any of the concepts.

Table 1: Set of materials corresponding to the concept of *negative feedback*

<table>
<tr><td colspan="2">Definition: Type of functioning that ensures that when the values of a variable depart from an optimal level, said departure activates processes that counteract the deviation.<br>Comprehension questions: (1) "In the present example, which variable departs from a desired level?", (2) "Which processes are activated by such departure?", and (3) "What effects do the activated process produce?".</td></tr>
<tr><td>Example 1 (Engineering)<br>The thermostat of a gas stove keeps its surroundings at a predefined temperature. When the room temperature raises over the preset temperature, the flame of the burner diminishes its power and the temperature of its surroundings begins to drop. When it drops below the preset temperature, the flame increases in power and the room temperature increases again.</td><td>Example 5 (Biology)<br>Environmental features affect the hydration of plants. When a plant starts to get dehydrated due to a drop in the humidity of the air, it responds to this environmental circumstance by orienting its leaves upward, thus reducing the surface that is exposed to the Sun. This in turn reduces evaporation, helping the plant to resume the normal degree of hydration of its tissues.</td></tr>
<tr><td>Example 2 (Engineering)<br>Water tanks have a valve that regulates the input of water from the network. The valve is connected to a ball float through a lever. When the height of the water drops, the weight of the float triggers the opening of the valve, and the tank gets refilled. Once the original height is achieved, the ball float lifts the lever and closes the valve, preventing the tank from overflowing.</td><td>Example 6 (Psychology)<br>In families where the parents do not control their aggressive impulses, a sibling may display psychological symptoms without an apparent cause. When violence between the parents begins to escalate, such sibling begins to display the pathological symptoms, capturing her parents' attention and thus reducing violence. When violence disappears, so does the pathological behavior.</td></tr>
<tr><td>Example 3 (Engineering)<br>The batteries from laptops and cell phones suffer some permanent damage when they get charged to its maximum real capacity. To prevent this from happening, advanced electronic devices cut the input of electricity when the battery charge surpasses 90% of its real capacity, and resume the input of electricity when the level of charge falls below said level.</td><td>Example 7 (Economics)<br>The price of hydrocarbons depends on a number of factors. When suppliers raise fossil fuels too much, consumers tend to use their vehicles less often, or to heat their homes less generously. This reduction in consumption leads to an increase of gas reserves by the suppliers. As this overstock is not profitable, suppliers tend react by lowering the price of their fuels.</td></tr>
<tr><td>Example 4 (Engineering)<br>Old torpedoes had an internal compass that recorded the initial direction at the moment it was launched. If the direction got altered by currents or other obstacles, the compass detected such change in direction, and steered the rudder opposite to the change. When the new direction matched that of the launch, the rudder got straight and the torpedo resumed a rectilinear trajectory.</td><td>Near-miss<br>When a meteor of astronomic dimensions impacts the Earth, it generates a cloud of ashes that blocks the sunlight, thus reducing the temperature of the planet's surface. Part del ocean near the Poles becomes frozen, therefore reflecting the solar radiation instead of absorbing it. This renders the planet even colder, affecting the life of animals and plants.</td></tr>
</table>

Table 2: Sample of how the seven examples of each set of materials were distributed across the two phases of the experiment

| | Learning Phase | Application Phase | | |
|---|---|---|---|---|
| Condition: | | Studied examples | Novel examples from learned domains | Novel examples from novel domains |
| Homogeneous | Examples #1, #2, and #3 from Engineering | Example#1 from Engineering | Example#4 from Engineering | Example from Economics |
| Varied | Example#1 from Engineering, plus examples from Biology and Psychology | Example#1 from Engineering | Example#4 from Engineering | Example from Economics |

For each of the learned concepts, the test included a studied example (i.e., the one that appeared in both conditions of the learning form), a novel example from a domain covered by the learned examples (i.e., the fourth of the examples of each set that belonged to the same domain), a novel example from a domain not covered by the learned examples, and a near miss (Table 2 displays a sample of how examples were assigned to phases and conditions). The test form also included five filler items, for a total of 25 trials. The order of presentation of the test items was counterbalanced by generating two versions of the learning form that featured the same items in two different random sequences. Participants took around 30 min to complete this second phase.

## Results

A mixed analysis of variance (ANOVA) with Type of test item (4 levels: studied examples, novel examples from studied domains, novel examples from non-studied domains and near misses) as within-subjects variable and Learning condition (2 levels: varied examples vs. homogeneous examples) as between-subjects variable revealed main effects of Type of test item, $F(3, 49) = 51.541$, $p < .001$ $\eta^2 = .759$ and of Learning condition $F(1, 51) = 4.189$, $p < .05$ $\eta^2 = .076$, but not a significant interaction $F(3, 49) = 1.135$, $p > .05$ $\eta^2 = .065$. Pairwise comparisons (Bonferroni) revealed that studied items were better classified than novel items taken from learned domains, and that these were in turn classified better than novel items from new domains (both $p_s < .001$). Participants' accuracy in classifying items from all three critical categories was higher than their accuracy in rejecting near-mises ($p < .001$, $p < .001$, and $p < .05$, respectively). Pairwise comparisons (Bonferroni) also revealed that participants in the varied examples condition outperformed participants in the homogeneous condition both in correctly classifying novel items from non-studied domains, and in correctly rejecting near misses (both $p_s < .001$), but neither in correctly classifying studied items nor in classifying new items from domains covered by the examples given for the corresponding concept (both $p_s > .1$). Figure 1 displays the mean classification scores across learning condition and type of test item.

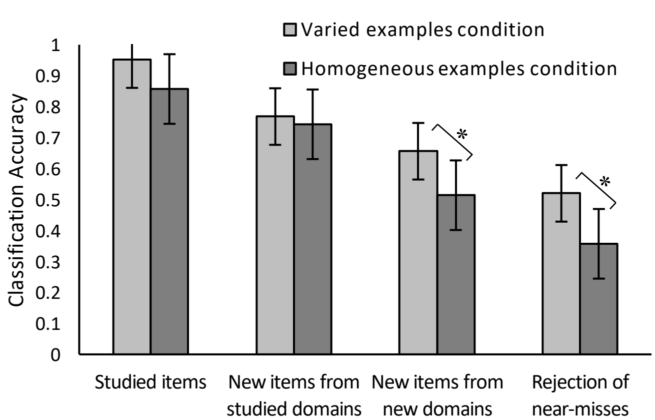


Figure 1: Mean classification scores across learning condition and type of test item.

## Discussion

The *problem of inert knowledge* (Whitehead, 1929) refers to the observation that learning often fails to be applied to contents and in contexts different from those in which it was acquired. Declarative concepts (abstract relational categories denoted by brief definitions) do not escape the problem of inert knowledge, since learners often fail to notice further exemplars not covered by the learning materials. Even though provided examples have proven superior than student-generated examples for concept application, classification performance after as many as 4 examples is still far from ceiling. Hence, more research was needed to determine how to select examples in ways that optimize concept application. The present study sought to contribute to this line of research by concentrating on one particular dimension along which examples can vary: their thematic domain. While there was some evidence that thematic variation aids the application of concepts pertaining to formal domains such as mathematics or statistics, evidence about its usefulness for science concepts of more empirical nature was scarce and negative (e.g., Corbalan, Kester & Van Merrienboer, 2009 system control condition; Day et al., 2010) presumably due to the added complexity involved in discerning between superficial and structural features of empirical phenomena. By employing five different system principles that cut across a wide range of thematic domains, we were the first to obtain robust evidence on the advantage of varied examples for cross-domain concept application.

The central finding of the present research was that coupling abstract conceptual definitions with thematically varied examples allowed participants to better recognize instances of the concept in domains not covered by the provided examples. This result is especially noteworthy given that the only available precedent for this kind of concepts was Day et al.'s (2010) study on negative vs. positive feedback, where the authors failed to obtain an advantage of thematically varied examples during a classification task. As we included negative feedback among our six system principles, we inspected the performance of our groups on this particular concept. Our data revealed that participants' performance in recognizing novel instances of negative feedback pertaining to domains not covered by the studied examples was significantly lower than their average performance on the other 4 concepts on this same measure, $M = 0.34$ $SD = 0.48$ vs. $M = 0.74$, $SD = 0.29$, $t(52) = 5.828$, $p < .001$, indicating that *negative feedback* might be particularly challenging for students. Even though the *n*s of our groups were not large enough to allow calculating Chi-square tests with sufficient power, an inspection of the data reveals a numerical advantage of the varied condition in extrapolating the concept to a novel domain (40% vs. 28,6%), suggesting that the advantage of example variation on this particular concept follows the general trend observed with the other concepts. Hence, our results help rectifying the pessimistic conclusions extracted by Day et al. out of a single concept, thus providing a robust demonstration that thematic variety alone (i.e., without asking participants to explicitly compare the examples) can aid cross-domain extrapolation of declarative concepts.

As compared to the above result, our finding that exposure to examples from three different domains is not inferior than exposure to three homogeneous examples for recognizing novel examples from the domains covered by the learned examples might seem at first sight as inconsequential. But it is not, especially if one considers that the association of, say, *negative feedback* to the domain of engineering would quite predictably be stronger after three engineering examples than after a single example from this domain. Therefore, it would have been unsurprising to obtain an inferior classification performance of the varied condition in recognizing new examples from studied domains. That said, it seems quite relieving to confirm that introducing variation across learning examples broadens the scope of application of the concept, but without paying the cost of a lesser ability to recognize further cases within the domains covered by the study examples.

In a similar vein, it could have been the case that achieving a more generalized representation of declarative concepts of the kind employed in the present study came at the cost of a diminished ability to reject near misses, especially when they belong to domains that have been covered by the examples. However, rejection of near misses proved to be more accurate in the varied condition, thus confirming that this instructional strategy can be safely employed to maximize distant transfer, without the danger of relaxing the criteria with which the learned category is applied to new exemplars.

Regarding mechanisms, several hypotheses could explain the observed effect of thematic variation on classification flexibility. One possibility is that the semantics of overly abstract notions such as "variable", "value", "agent", "process" or "resource" are very difficult to mentally represent, and therefore the examples serve to disambiguate such terms. On this account, as the elements that fill the role of "variable" in concepts like negative feedback tend to be very similar across homogeneous examples of the definition, their conceptual intersection still preserves domain-specific features, therefore conspiring against a flexible application of the concept. While this explanation is perhaps the most appealing, it would seem to be at odds with traditional results from the analogical retrieval literature, which convergently demonstrate that schema induction rarely occurs in the absence of an explicit comparison between the analogs (e.g., Gentner et al., 2003). It is at least conceivable that, as opposed to idiosyncratic structural isomorphs such as Gick and Hoyoak's (1983) stories, members of lexicalized categories like the system principles here employed invite spontaneous comparisons between category members, especially when they are the explicit targets of instruction. However, previous studies like Goldwater & Gentner (2015) suggest that even in these relatively advantageous cases spontaneous comparison among exemplars is unlikely.

Another possibility is that far from eliciting a more general schema, examples of declarative concepts are themselves being recruited when assessing a candidate for inclusion into the category. Given that structural isomorphs are seldom retrieved in the absence of a concomitant degree of superficial similarity (Keane, 1987, Kurtz & Loewenstein, 2007, see Trench & Minervino, 2017 for a review), a selection of thematically varied examples might increase the odds that at least one of the learned examples maintains a degree of surface similarity with the candidate being processed.

Albeit uninspiring, yet another possibility is that the observed advantage of varied examples originates neither in a more abstract schema nor on a set of examples whose superficial features achieve more "coverage" across the space of potential base analogs, but rather in coupling the received definition with a "domain tag" (Ripoll, 1998) that somehow specifies the thematic range of situations to which the abstract relations conveyed by the definitions may apply. To exemplify, the presentation of varied examples could have neutralized a preexisting—perhaps even conscious—assumption that *inoculation* refers to phenomena related to viruses or bacteria.

While carefully crafted materials could probably be developed to decide between the schema-based and the exemplar-based accounts of the effect of thematic variation, the truth is that the conditions that favor one or the other processes seem to be tightly correlated in the natural world. From our own experience in struggling to construct the present sets of materials, it became clear to us that the possibility of instantiating even concepts as broad as our system principles in further and further unrelated domains reaches a ceiling quite abruptly, typically after 4 or 5 domains. Beyond that limit, the elements in these new instantiations inevitably begin to maintain surface similarities with the elements of the examples taken from the previous domains, even when the domains themselves cannot be said to match. Hence, meeting the conditions for schema induction also yields the broader coverage of base situations that sets out the conditions for analogical transfer. This is good news for the prospects of *transfer*, a research area often unaccustomed to receiving auspicious results.

## Acknowledgments

This work was supported by Grant PIP 11220200103018 from the National Council for Scientific and Technical Research (CONICET) and by Grants PICT 2017-0853, PICT 2019-02542 and PICT 2019-3268 from the National Agency for Scientific and Technological Research (ANPCyT).

## References

Balch, W. R. (2005). Elaborations of introductory psychology terms: effects on test performance and subjective ratings. *Teaching of Psychology, 32*, 29–34.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin, 128*, 612–637

Ben-Zeev, T., & Star, J. R. (2001). Spurious correlations in mathematical thinking. *Cognition and Instruction, 19*, 253–275.

Braithwaite, D. W., & Goldstone, R. L. (2015). Effects of variation and prior knowledge on abstract concept learning, *Cognition and Instruction, 33*, 226-256.

Chang, N. M. (2006). *Learning to discriminate and generalize through problem comparisons*. Doctoral Dissertation, Carnegie Mellon University, Pittsburgh, PA.

Chang, N. M., Koedinger, K. R., & Lovett, M. C. (2003). Learning spurious correlations instead of deeper relations. In R. Alterman & D. Kirsch (Eds.), *Proceedings of the 25th annual conference of the Cognitive Science Society* (pp. 228–233). Boston, MA: Cognitive Science Society.

Chen, Z., & Mo, L. (2004). Schema induction in problem solving: A multidimensional analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 583–600.

Corbalan, G., Kester, L., & van Merrienboer, J. J. G. (2009). Combining shared control with variability over surface features: Effects on transfer test performance and task involvement. *Computers in Human Behavior*, *25*, 290–298.

Day, S. B., & Goldstone, R. L. (2012). The import of knowledge export: Connecting findings and theories of transfer of learning. *Educational Psychologist*, *47*, 153–176.

Day, S. B., Goldstone, R. L., & Hills, T. (2010). *The effects of similarity and individual differences on comparison and transfer*. In S. Ohlsson & R. Catrambone (Eds.), Proceedings of the 32nd annual meeting of the cognitive science society (pp. 465–470). Portland, OR: Cognitive Science Society.

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology*, *95*, 393–405.

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

Goldstone, R. L., & Sakamoto, Y. (2003). The transfer of abstract principles governing complex adaptive systems. *Cognitive Psychology, 46*, 414-466

Goldstone, R. L., & Wilensky, U. (2008). Promoting transfer by grounding complex systems principles. *Journal of the Learning Sciences, 17,* 465-516.

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.

Jamrozik, A., & Gentner, D. (2020). Relational labeling unlocks inert knowledge. *Cognition*, 196, 104146.

Keane, M. T. (1987). On retrieving analogues when solving problems. *Quarterly Journal of Experimental Psychology, 39*, 29–41.

Kurtz, K., & Loewenstein, J. (2007). Converging on a new role for analogy in problem solving and retrieval: When two problems are better than one. *Memory & Cognition, 35*, 334-341.

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

Merrill, M. D., Tennyson, R. D., & Posey, L. O. (1992). *Teaching concepts: An instructional design guide* (2nd Edition). Englewood Cliffs, NJ: Educational Technology Publications.

Minervino, R., Olguín, V., & Trench, M. (2017). Promoting interdomain analogical transfer: When creating a problem helps to solve a problem. *Memory & Cognition*, *45*, 221-232.

Paas, F. G. W. C., & Van Merrienboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology*, *86*, 122–133.

Quilici, J. L., & Mayer, R. E. (1996). Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology*, *88*, 144–161.

Quilici, J. L., & Mayer, R. E. (2002). Teaching students to recognize structural similarities between statistics word problems. *Applied Cognitive Psychology*, *16*, 325–342.

Rawson, K. A., Thomas, R. C., & Jacoby, L. L. (2015). The power of examples: Illustrative examples enhance conceptual learning of declarative concepts. *Educational Psychology Review, 27*, 483-504.

Ripoll, T. (1998) Why this makes me think of that, *Thinking & Reasoning*, *4*, 15-43.

Ross, B. H. (1987). This is like that: The use of earlier problems and the separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 629–639.

Trench, M., & Minervino, R. A. (2017). Cracking the problem of inert knowledge: Portable strategies to access distant analogs from memory. In B. H. Ross (Ed.), The psychology of learning and motivation (Vol. 66, pp. 1–41). San Diego, CA: Academic Press.

Whitehead, A. N. (1929): *The Aims of Education and Other Essays*. New York: The Free Press.

Zamary, A. (2019). *Evaluating the dual-route and recruitment hypotheses: Utilizing both definitions and examples for supporting declarative concept application.* Doctoral Dissertation, Kent State University, Kent, OH.

Zamary, A., & Rawson, K. A. (2018a). Which technique is most effective for learning declarative concepts-provided examples, generated examples, or both? *Educational Psychology Review, 30*, 275-301.

Zamary, A., & Rawson, K. A. (2018b). Are provided examples or faded examples more effective for declarative concept learning? *Educational Psychology Review, 30*, 1167-1197.

Zhang, D. (2019). *Remindings in declarative concept learning*. Masters Dissertation, University of Arizona, Tucson, AZ.