Available online at www.sciencedirect.com
SCIENCE DIRECT®
ACADEMIC PRESS
Cognitive Psychology 46 (2003) 414–466
Cognitive Psychology
www.elsevier.com/locate/cogpsych

The transfer of abstract principles governing complex adaptive systems

Robert L. Goldstone[a,*] and Yasuaki Sakamoto[b]

[a] *Indiana University, Bloomington, IN 47408, USA*
[b] *University of Texas at Austin, USA*

Accepted 1 August 2002

Abstract

Four experiments explored participants' understanding of the abstract principles governing computer simulations of complex adaptive systems. Experiments 1, 2, and 3 showed better transfer of abstract principles across simulations that were relatively dissimilar, and that this effect was due to participants who performed relatively poorly on the initial simulation. In Experiment 4, participants showed better abstract understanding of a simulation when it was depicted with concrete rather than idealized graphical elements. However, for poor performers, the idealized version of the simulation transferred better to a new simulation governed by the same abstraction. The results are interpreted in terms of competition between abstract and concrete construals of the simulations. Individuals prone toward concrete construals tend to overlook abstractions when concrete properties or superficial similarities are salient.

1. Introduction

Scientific understanding frequently involves comprehending a system at an abstract rather than superficial level. Biology teachers want their students to understand the genetic mechanisms underlying heredity, not simply how pea plants look. Physics teachers want their students to understand fundamental laws of physics such as conservation of energy, not simply how a particular spring uncoils when

* Corresponding author.
E-mail address: rgoldsto@indiana.edu (R.L. Goldstone).
URL: http://cognitrn.psych.indiana.edu/.

0010-0285/03/$ - see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S0010-0285(02)00519-4

weighted down (Chi, Feltovich, & Glaser, 1981). This focus on acquiring abstract principles is well justified. Science often progresses by researchers finding deep principles shared by superficially dissimilar phenomena, and by describing situations in terms of mathematical or formal abstractions. Finding biological laws that govern the appearance of both snails and humans (Darwin, 1859), physical laws that govern both electromagnetic and gravitational acceleration (Einstein, 1989), and psychological laws that underlie transfer of learning across species and stimuli (Shepard, 1987) are undeniably important enterprises. Transcending superficial appearances to extract deep principles is as critical to science as it is difficult to achieve. These deep principles are often called "schemata," and are important for high-level cognition because once they have been acquired, they can be applied to a wide variety of subsequent problems (Holyoak, 1984; National Research Council, 1999; Novick & Holyoak, 1991; Rumelhart, 1980).

The currently reported research explores factors that facilitate the transfer of an abstract scientific principle from one domain to another. If abstracting deep principles that cut across different domains is frequently valuable (see Anderson, Reder, & Simon, 1996 and Barnett & Ceci, 2002 for defenses of this assumption), then it is likewise valuable to find ways to promote this abstraction. In the present experiments, participants' success in acquiring deep principles is measured by both direct and indirect methods. Participants are directly tested by multiple-choice quizzes, and appreciation of an abstract principle is indirectly measured by the degree of transfer between superficially dissimilar domains that are governed by the same principle. With these measures, experiments are conducted to explore whether increasing the superficial similarity between domains promotes transfer of the abstract principle shared between them (Experiments 1–3), and whether perceptually rich or impoverished presentations better promote transfer of an abstract principle (Experiment 4). A central theme throughout these experiments concerns the relation between the superficial, concrete details through which a phenomenon is presented, and the ability of people to extract the deeper scientific principles underlying the phenomenon.

2. Perceptual richness in computer simulations

A desire to convey abstract scientific principles does not necessarily entail that the favored instructional strategy will involve direct presentation of these abstractions in a mathematical or verbal form. In fact, a wealth of research indicates that graphic, perceptually salient displays are more effective than purely statistical or verbal information. Decision makers are more strongly affected by vivid information than either abstract statistical information (Nisbett & Ross, 1980) or pallid information (Reyes, Thompson, & Bower, 1980). Even when abstract understanding is desired, concrete representations are often advantageous (Barsalou, 1999; Cheng, 2002; Goldstone, 1994a, 1994b; Goldstone & Barsalou, 1998). For example, scientific (Gentner & Stevens, 1983) and logical (Johnson-Laird, 1983) situations are often successfully understood by constructing working mental models that concretely instantiate elements of the world to be explained. Barsalou (1999) has described the importance of "perceptual simulations" that use perceptual processes such as mental imagery to simulate even abstract concepts such as negation and truth.

If a concrete picture is worth a thousand words, then dynamically changing and interactive pictures should be worth even more. This is part of the motivation behind the increasing use of computer-based simulations for teaching scientific concepts. A computer simulation explicates scientific concepts by creating simplified working models that are typically under parametric control by the simulation's user (Miller, Lehman, & Koedinger, 1999; Resnick, 1994; Schank & Farrel, 1988). Computer simulations have been shown to confer a number of advantages over traditional educational practices. First, they provide a perceptual grounding for concepts that might otherwise be too abstract to readily be comprehended. Second, they promote an active, hands-on, problem-solving stance by learners that, in turn, often fosters a deep understanding of a phenomenon (National Research Council, 1999). Third, they provide effective exposure to experimentation skills that involve a cycle of hypothesis formation, testing, evaluation, and revision (White, 1993; White & Fredericksen, 1998).

In recent years, educational practices involving the use of computer simulations have grown dramatically, and show little sign of abating. However, most simulations are designed without exploring and testing design choices that could have a major impact on learning (for some exceptions, see Jackson, Stratford, Krajcik, & Soloway, 1996; Klahr & Carver, 1988; Miller et al., 1999). Even when researchers empirically analyze the educational outcomes derived from a particular computer simulation, they rarely explore systematic differences in the simulation to try to optimize its educational impact or to assess what elements of the simulation were critical for imparting the educational benefit. Cognitive psychology and formal experimentation have a major role to play in improving the pedagogical impact of computer simulations.

One of the particular design issues addressed in the current experiments concerns how much superficial detail to include in a simulation. A high level of detail, produced by a realistic rendering of objects within the simulation, may benefit learners by making a phenomenon concrete and by increasing the similarity between the simulation and real-world situations (DiFonzo, Hantula, & Bordia, 1998). In fact, most research in virtual reality has as an explicit goal the realistic mimicking of real-world phenomena (Grady, 1998; Heim, 2000). Highly realistic computer simulations may be entertaining, evocative, and impressive, but it is not yet clear that they are superior at teaching people about abstract scientific principles. Other researchers have argued that simplified, relatively idealized representations are useful for distilling a situation to its essence (Gianutsos, 1994). An impressive empirical case for the value of less realistic representations comes from DeLoache's research on children's use of physical models as representations (DeLoache, 1991, 1995; DeLoache & Burns, 1994; DeLoache & Marzolf, 1992). In her standard paradigm, a child around the age of 2.5 years is shown a model of a room, the child watches as a miniature toy is hidden behind or under a miniature item of furniture in the model, and she or he is told that a larger version of the toy is hidden with the corresponding piece of furniture in the room. Children around the age of 2.5 years were better able to use the model to find the toy in the actual room when the model was a two-dimensional picture rather than a three-dimensional scale model (DeLoache, 1991; DeLoache & Marzolf, 1992). DeLoache (1995) explains this result in terms of the difficulty in understanding an object as both a concrete, physical thing, and as a symbol standing for something else. As an object's concrete, physical properties become more salient, its ability to serve as a symbol decreases. A two-dimensional picture serves as a better symbol because its concrete properties are less salient. Consistent with this account, when the physical salience of a model is reduced by placing it behind a window, or by limiting contact with it, it serves as a more effective symbol, and children are better able to find the toy based on information from the model.

DeLoache's work might be extended to argue that a vividly concrete representation interferes with the extraction of abstract principles from the representation. In fact, Uttal, Liu, and DeLoache (1999) argue that concreteness is not always beneficial. Although concreteness can help young children detect symbolic relations, it can make it more difficult for them to comprehend the abstract concepts represented by the concrete object. Another example of this is found in transfer among word problems in mathematics. Bassok and Holyoak (1989) examined transfer between isomorphic domains of arithmetic progression in algebra and constant acceleration in physics. They found much greater positive transfer from the algebra problems to the physics problems than vice versa. They interpreted their results as showing that when abstract mathematics is easily isolated from the content-specific cover story in which it occurs, it transfers widely to different situations (for a modified account and extension of this result, see Bassok & Olseth, 1995). Likewise, children (Ratterman & Gentner, 1998) and adults (Markman & Gentner, 1993) are less likely to respond on the basis of abstract relations among objects in a scene and more likely to respond on the basis of superficial object attributes as the richness of the objects in the scene increases. As with the DeLoache studies, when the concrete manifestation of an abstraction is difficult to ignore, it adversely impacts responding on the basis of abstractions.

3. Complex adaptive systems

The currently studied computer simulations involve scientific principles underlying complex adaptive systems. Complex adaptive systems are systems made up of many units (often times called agents), whose simple interactions give rise to higher-order emergent behavior. Typically, the units all obey the same simple rules that control how they interact, but because of these interactions, the units may become specialized and individualized. Despite the lack of a centralized control, leader, recipe, or instruction set, these systems self-organize themselves into a state of global organization (Resnick, 1994). Many scientific domains can be explained by the formalisms of complex adaptive systems, including the foraging behavior of ants, the development of the human nervous system, the growth of cities, growth in the world wide web, the perception of apparent motion, mammalian skin patterns, pine cone seed configurations, and the shape of shells (Ball, 1999; Casti, 1994; Flake, 1998). One reason for taking complex adaptive systems as the domain of inquiry is because of its intrinsic importance to cognitive science, and the increasing need to understand these systems.

A second reason for considering complex adaptive systems is because of the applicability of their scientific principles across many domains. An underlying assumption of complex adaptive systems research is that widely different systems can be explained by the same mathematical formalisms. The same reaction–diffusion equations that explain how spots develop on a leopard can be used to explain how regional pockets of similar political or religious beliefs develop in a country. The same differential equations that govern the dynamics of populations of predators and preys can explain fluctuations in businesses and oscillating chemical reactions (Ball, 1999). Sand piles, earthquakes, and human memory for temporal intervals can all be understood as systems that naturally adapt to a point of self-organized criticality governed by a $1/f$ power spectrum (Bak, 1996; Gilden, Thornton, & Mallon, 1995). This cross-domain applicability of complex adaptive system principles is valuable for the current psychological investigation because it allows for a natural examination of the extent of transfer of an abstract principle to different domains. Instead of creating word problems with different cover stories or abstract schemata that can be instantiated with different insight problems, domains can be selected that naturally and intrinsically instantiate principles of complex adaptive systems. The advantages are that the systems are inherently, rather than arbitrarily, connected to their abstract principles (see Bassok, 1996; Bassok, Chase, & Martin, 1998), the principles have external validity in that they are of authentic scientific interest, and comprehending the principles is challenging even for college students. Complex adaptive systems offer a unique laboratory for exploring abstract and concrete thought because the domains of inquiry embody abstract formalisms. To understand such a system requires both knowing its details and knowing how the details embody a general principle.

4. Analogical transfer across simulations governed by the same principle

We are interested in what allows the deep principle that is instantiated by a simulation to be transferred to another domain that uses the same principle. This line of inquiry falls squarely in the field of analogical reasoning. We are interested in analogical reasoning as a way of diagnosing the depth of learning that results from exposure to a simulation. That is, we will take it as evidence that experience with a simulation allows a learner to understand the deep principle underlying it if the learner shows better understanding of a subsequent system governed by the same principle than a control learner who was not exposed to the same principle. Measuring understanding of a deep principle in this way is useful because of its external validity. In most educational settings, material is given to students in the hope that they will not simply master the material in the classroom setting, but will apply the principles learned to new, real-world situations (Barnett & Ceci, 2002; Reeves & Weisberg, 1994). One of the major issues that researchers in analogical reasoning have explored is the relation between the superficial (surface-level) similarity between domains and transfer between them. Several researchers have found stronger transfer from one domain to another if the domains are superficially similar to each other. In solving a word problem, participants are highly influenced by the previous solution of a problem if it involves the same superficial "cover story" (e.g., both problems involve golf) (Ross, 1987). Abstractly related problems are more likely to be accessed when trying to solve a problem if the analogous problems also have superficial resemblances to the unsolved problem (Gentner, 1989; Gentner & Toupin, 1986; Holyoak & Koh, 1987; Keane, 1987; Ross, 1984, 1990; Reed, 1999). Once superficial similarity prompts two situations to be compared, abstract information from one situation can be applied to the other situation (Gentner, 1989; Novick & Holyoak, 1991; Ross, Perkins, & Tenpenny, 1990). One of the largest benefits of increasing the superficial similarity between problems is that it increases the likelihood that a person will be reminded of the previously solved problem when considering a new problem (Gick & Holyoak, 1980, 1983; Holyoak & Koh, 1987; Ross, 1987, 1990).

Superficial similarity between problems has been manipulated in a number of ways. Sometimes it is manipulated by changing the "cover story" that exemplifies a deep principle. In these cases, increasing the similarity between cover stories generally increases transfer between them (Keane, 1987; Ross, 1984, 1989). It can also be manipulated by varying the superficial similarity between objects across domains. Generally speaking, increasing the similarity of objects that play analogous roles in their domains increases transfer, even if the domains are dissimilar (Gentner & Toupin, 1986; Holyoak & Koh, 1987; Ross, 1989). In the absence of superficial similarity that connects dissimilar but analogous problems, transfer is typically poor unless hints are provided about the relevance of the old problem while solving the new problem (Gick & Holyoak, 1980, 1983).

Work with children has also indicated that providing superficial similarities between situations can allow more abstract commonalities to be appreciated. Three-year-old children can match sets on the basis of numerosity better when the sets are made of the same objects (Mix, 1999). Roughly put, 3-year-old children can respond that XX is more like XX than XXX even when they do not reliably respond that XX is more like OO than OOO. Similarly, Kotovsky and Gentner (1996) found that 4-year-old children are better able to appreciate an abstract symmetry relation shared between two situations when they are first trained on literal similarity comparisons in which the situations share both superficial and abstract commonalities. They argue that close literal similarity matches facilitate subsequent analogical matches by causing situations to be placed in correspondence, and once placed in correspondence, the abstract relations that they share are highlighted (see also Gentner & Wolff, 2000). Apparently, superficial similarities, when they promote the same correspondences between scenarios as do abstract similarities, can facilitate the extraction of these abstract similarities by suggesting appropriate correspondences between domains.

Although the bulk of research has suggested that superficial similarities between domains promote comparison between the domains and hence the extraction of shared abstractions, there are some studies that conflict with this conclusion. To the extent that there is competition between superficial and abstract construals of a situation, increasing superficial similarity may highlight superficial construals, thereby impeding the processing of abstractions. For example, Goldstone, Medin, and Gentner (1991) showed participants a standard scene and different pairs of alternatives, and asked them to judge which of the alternatives was more similar to the standard. One of the alternatives was superficially more similar to the standard, and the other alternative was more abstractly similar. In Fig. 1, the standard "X star X" shares a superficial "X" match with "X circle square" and shares the more abstract relational similarity of "same-shapes on the ends" with "square circle square." When the superficial similarity of the scenes was relatively low due to the star/circle mismatch, people tended to select the abstract alternative as most similar to the standard T in Fig. 1. This tendency was significantly reduced when both of the alternatives had higher superficial similarity to the standard, produced by replacing each of the circles in the alternatives by stars, as shown in the lower panel of Fig. 1. Goldstone et al. argue that when superficial similarity among compared scenes is high,

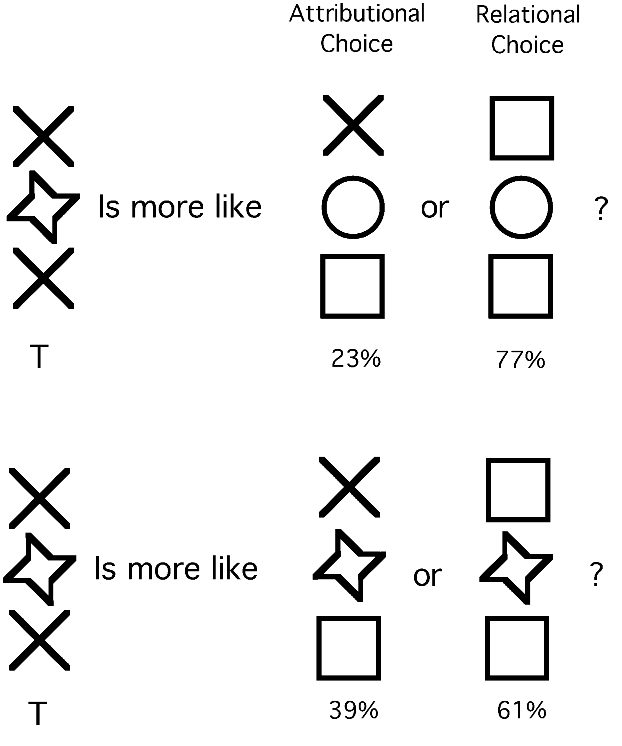

Fig. 1. A sample stimulus set from Goldstone et al. (1991). Participants were asked to decide which of two sets of objects was more similar to the object set T. When the two choices had relatively few superficial properties in common with T (top panel), participants showed a strong tendency to choose the set with more abstract, relational properties in common with T. When the two choices had more superficial properties in common with T (bottom panel), participants selected the set with concrete, superficial properties in common with T more often than they did in the top panel.

people will tend to use this level when analyzing scenes, consequently missing abstract similarities more often than when superficial similarity is lower.

Abstractions shared by scenarios are sometimes better appreciated when the scenarios are superficially dissimilar rather than similar. For example, Gick and Holyoak (1983, Experiment 6) gave their participants two abstractly related problems from the same domain (military or fire-fighting) or from different domains. Although there was no difference between similar and dissimilar problems in facilitating transfer to another analogous problem, the dissimilar problems did result in participants producing better written schemas that captured the abstract similarities between the problems. When two dissimilar problems are compared, they share relatively little in common other than an abstract commonality, and thus this abstraction may be highlighted because of its uniqueness. In contrast, two similar problems share both superficial and abstract commonalities, and the abstraction may not be emphasized (Medin & Ross, 1989). If comparing two situations emphasizes their commonalities, then abstract commonalities will receive the most attention when the situations are relatively dissimilar.

In summary, there are reasons for thinking that superficial similarities may either promote or interfere with the apprehension of abstract commonalities. On the one hand, superficial similarities may promote the retrieval of one situation when cued with the other, and may promote an alignment process that reveals their abstract similarities (Gentner, 1983; Gentner & Wolff, 2000; Goldstone, 1994a, 1994b; Goldstone & Medin, 1994; Markman & Gentner, 1993). On the other hand, salient superficial similarities may dissuade people from looking for similarities at a deeper level, or may decrease the uniqueness and therefore salience of abstract similarities.

5. Individual differences in transfer

The preceding discussion of the role of similarity in analogical transfer needs to be qualified by potential individual differences. Individual differences related to the level of a participant's manifest understanding are important for interpreting our current experimental results. An often observed pattern is that relatively good performers are less influenced by surface-level representations than poorer performers, and are more influenced by abstract representations (Chi et al., 1981; Miller & Stigler, 1991; Schiano, Cooper, Glaser, & Zhang, 1989). People exhibiting unusually good comprehension of a situation typically have particularly good understandings of the deep, fundamental principles even though their memory for details may not be especially good (Voss, Greene, Post, & Penner, 1984). Campione (1985) found that good comprehenders show better transfer to new problems than poor comprehenders even when the groups are equated in terms of their performance on an original set of problems. Furthermore, the advantage for good comprehenders increased as the dissimilarity between the original and transfer problems increased. Thus, poor performers' knowledge seems to be relatively tightly tied to the original materials and does not generalize well to abstractly related materials. In addition to these rather coarse characterizations of individual differences, computational models of individuals' understandings of geometric problems can be effective in predicting their transfer performance (Lovett & Anderson, 1996).

Even granting that poor comprehenders of a domain are relatively reliant on the concrete rather than abstract aspects of a situation, conflicting predictions are still possible for the interaction between domain similarity and learner's comprehension on transfer between domains. One might predict that poor comprehenders will be particularly hurt by decreasing the superficial similarity between two abstractly related domains. Poor comprehenders may rely on superficial similarities to give them the notion to connect the two domains (Holyoak & Koh, 1987). Once the two domains are connected by a reminding, an earlier learned principle may be applied to the new domain (Ross, 1984; Ross et al., 1990). In this manner, the superficial similarities allow a poor comprehender to bootstrap their way to a more sophisticated understanding (Mix, 1999; Ratterman & Gentner, 1998).

The other possible prediction is that poor comprehenders are particularly hurt by increasing the superficial similarity between abstractly related domains. Poor comprehenders are at particular risk for interpreting a situation in an overly concrete manner. If two situations are superficially similar, then the generalization that the poor comprehender forms may include these superficial similarities and consequently may not emphasize their shared abstract principle (Hummel & Holyoak, 1997; Medin & Ross, 1989). The two predictions thus described share the premise that poor comprehenders are more influenced by superficial similarities than are good comprehenders, but differ in whether the influence of superficial similarities is thought to promote or hurt abstraction-based transfer (see the previous section). In essence, the difference between these predictions comes down to whether superficial similarities are seen as clues to deeper similarities (Goldstone & Barsalou, 1998; Medin & Ortony, 1989) or distractions that interfere with the processing of deeper similarities (Goldstone et al., 1991).

6. The current inquiries

The experiments to be reported investigate the role of concrete, perceptual information in interfering with or promoting the apprehension of abstract principles from computer simulations, and how this is modulated by individual differences. We investigated whether superficial similarity across domains promotes or interferes with transfer (Experiments 1–3), and whether perceptually rich or impoverished displays promote greater transfer (Experiment 4).

7. Experiment 1

Experiment 1 uses transfer between simulations to explore the role of superficial similarity in promoting such transfer. As described in Section 1, a considerable body of research has shown beneficial effects of superficial similarity (e.g., similarity in the cover stories used for algebraic word problems or semantic similarity between corresponding objects in insight problems) on analogical reminding and use. On the other hand, there was also some evidence that high degrees of superficial similarity may conceal more abstract commonalities. Experiment 1 explores transfer between abstractly related simulations, manipulating the superficial similarity between elements that play corresponding roles across the simulations.

In preparation for Experiment 1, a pilot experiment was conducted to ensure that transfer across abstractly related simulations could be found. In this pilot experiment, participants were given active experience with two simulations selected from a set of four. Two of the four simulations were governed by the principle of simulated annealing (the simulations used in Experiments 1–3), and the other two were governed by competitive specialization (the simulations used in Experiment 4). For one group of participants, the explored simulations were governed by the same principle, and for the other participants they were governed by different principles. Abstract transfer was measured by multiple choice quiz performance probing understanding of the second simulation's governing principle. We found significantly better quiz performance on the second simulation when it was preceded by an abstractly related rather than unrelated simulation during training. Although spontaneous transfer is not always found across analogically related problems (Gick & Holyoak, 1980, 1983), the pilot experiment, as well as the currently reported experiments, had several design aspects that have been empirically linked with abstract transfer. First, participants were engaged in active problem solving during the first simulation (Needham & Begg, 1991). Second, the physical (Barnett & Ceci, 2002) and temporal contexts of the two simulations were very close. Third, abstract descriptions of the rules governing the simulations were given to participants (Catrambone & Holyoak, 1989; Gick & Holyoak, 1983). Fourth, the abstract principles were given dynamic, perceptually grounded manifestations (Pedone, Hummel, & Holyoak, 2001; Reed, Cooke, & Jazo, 2002; White, 1993). Given that better transfer was found between related than unrelated simulations, in Experiment 1 we use the same simulations to explore how superficial similarity and individual differences affect the abstract transfer found in this pilot experiment.

The simulations used in Experiment 1 are two instantiations of the scientific principle of simulated annealing. Although the simulations will be briefly described here, this description is a poor substitute for dynamically exploring the simulations, which can be accessed via the web at http://cognitrn.psych.indiana.edu/rgoldsto/complex. Simulated annealing refers to a search technique that makes use of randomness in order to find optimal solutions to a problem (Kirkpatrick, Gelatt, & Vecchi, 1983). The notion of simulated annealing is to gradually reduce the randomness in a system. Early on, randomness helps the system sample different candidate solution spaces. Later on, stability helps the system settle down into a single strong solution. In computational applications of simulated annealing, the probability that a system will move from a state with energy $E1$ to a state with energy $E2$ is set to $e^{-(E2-E1)/kT}$, where a low E value means a good solution (energy is measuring discord) and T is the temperature or randomness in the system. If temperature is very low, then the system will almost always move in the direction of lower E. However, if T is high, then the system may move toward the state with a higher E. Although it may seem counter-productive to increase the energy of the system if one wants to have the least energy possible, it helps the system in avoiding local minima. Local minima are states such that the current E value is lower than all E values that are obtained by neighboring states, yet E is not the lowest possible value that the system could obtain. In order to find global rather than local minima, it is often necessary to temporarily increase the energy in the system.

7.1. Dropping balls

The first domain involving simulated annealing is balls dropping on a user-drawn landscape, as shown in Figs. 2 and 3. Having students imagine a ball rolling with some randomness on a landscape is one of the most commonly used analogies for teaching simulated annealing. Small red balls fall according to three rules: (1) A ball will tend to fall downward due to gravity. (2) A ball also moves with a user-controlled degree of randomness (due to chance winds). (3) If a ball's movement would place it on a user-drawn green patch (the landscape), then it does not move. Learners are given the general goal of developing an automatic strategy that will cause the balls to fall to the lowest region of the landscape they draw. The learner can control several aspects of the simulation by manipulating buttons, sliders, and the cursor. Starting with the configuration of balls and the user-drawn landscape in Fig. 2,

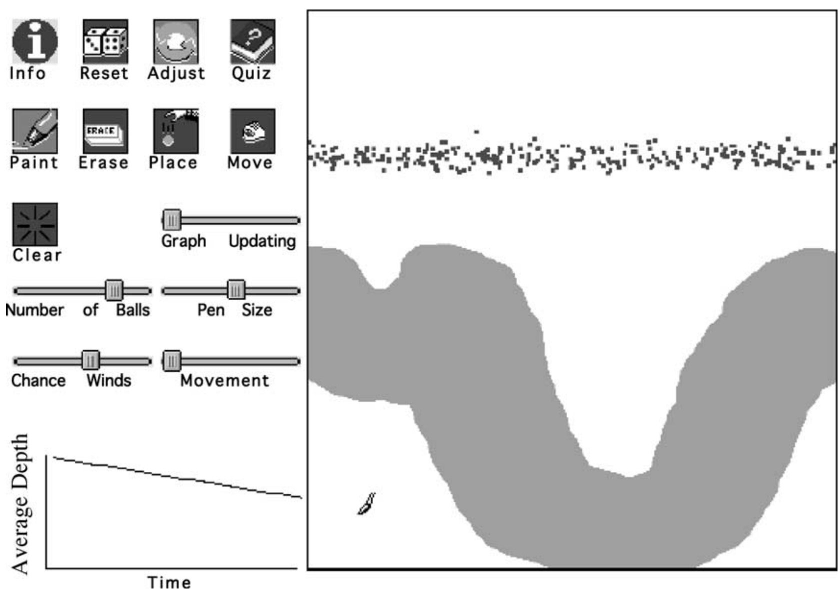

Fig. 2. A screen-dump from the "Ball Drop" simulation before the balls have completely dropped onto the landscape. User-controlled buttons and sliders, and the continuously updating graph, are shown on the left side. In the right window is a dynamically changing environment in which balls are dropping, landscapes are drawn and altered, and balls are selected and moved.

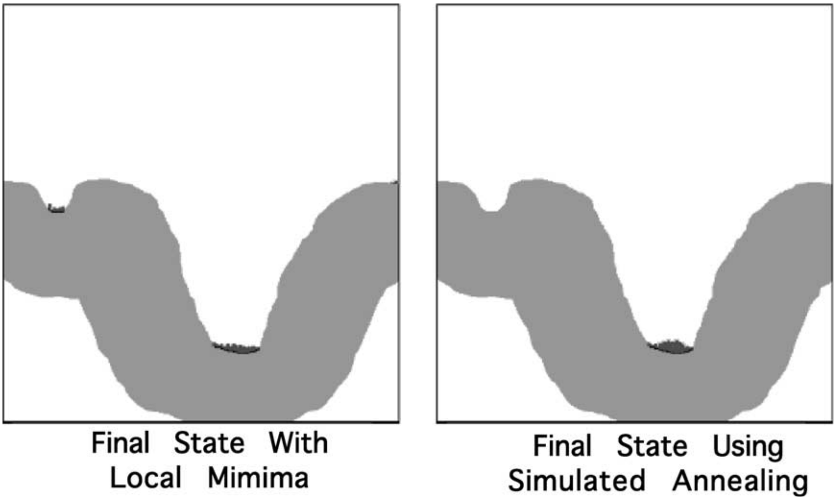

Fig. 3. Two possible final configurations for the "Ball Drop" simulation. If balls drop without much randomness added to their movements, then the final configuration of balls will typically show several local minima. A ball is in a local minima if its location is lower than neighboring locations, but is not the lowest location for the whole landscape. If the amount of randomness is gradually reduced as specified by a simulated annealing method, then all of the balls will occupy the lowest position on the landscape.

two possible end states are shown in Fig. 3. If the amount of user-controlled randomness is never very large, or if the amount of randomness is reduced too quickly, then local minima are likely to arise. A local minimum occurs if a ball falls to rest in a valley that is not the deepest valley of the landscape. If randomness is gradually reduced and thus consistent with simulated annealing, then all of the balls will eventually come to rest at the lowest spot on the landscape.

7.2. *Path finder*

The second example of simulated annealing involves finding a pathway around obstacles. The pathway ideally connects two fixed blue points at the top and bottom of the screen, avoids the green obstacles, and is as short as possible. In traditional artificial intelligence, the search for a pathway through a maze is typically viewed as a process of adding segments to a pathway and backtracking when dead-ends are found. The alternative method pursued here is to have simple agents locally influence each other's positions. They globally form a path even though no agent by itself represents an entire solution. This is accomplished by having a set of small red balls follow three rules: (1) Each ball is assigned two neighbors, making a set of balls arranged from first to last. One of the fixed blue points is the neighbor of the first ball. The other blue point is the neighbor of the last ball. (2) Each ball moves toward each of its two neighbors and also moves with a user-controlled amount of randomness. (3) If the location to which a ball would move is painted green, then it does not move. The buttons and parameters are similar to those used in "Dropping

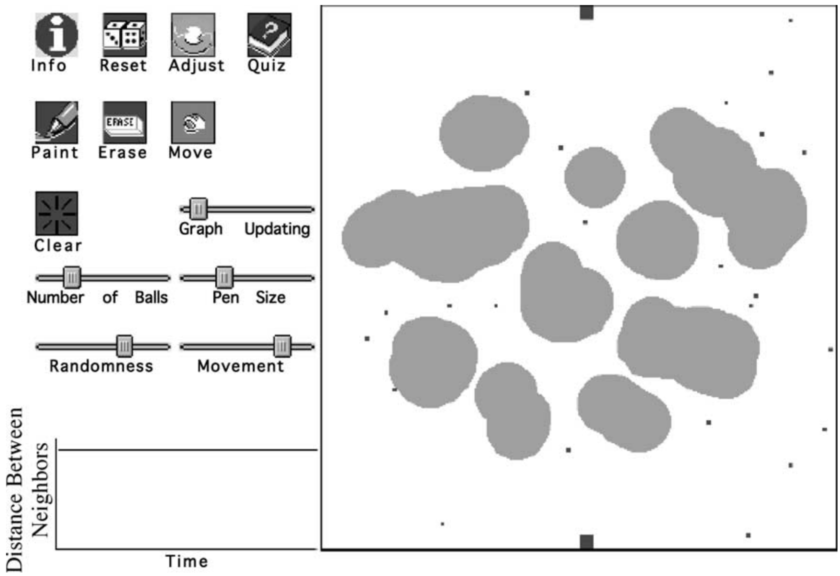

Fig. 4. A screen-dump from the simulation "Path finder." In this initial configuration, balls are randomly positioned on the screen. They move toward their pre-specified two neighbors, unless the movement will place them on top of a green patch.

balls" and are shown in Fig. 4. Starting with the initially random configuration in Fig. 4, two possible final configurations of the balls are shown in Fig. 5. In the configuration on the left, each of the balls is as close as it can get to its two neighbors without travelling through a green region. The configuration does not indicate a strong pathway between the fixed points and is typical of the kind of pattern that is found when the balls do not move with sufficient randomness or when the randomness is reduced too quickly. These "knots" cannot be avoided if the balls do not have some randomness that allows them to break out of arrangements that place them as close to their neighbors as possible given the constraints of Rule 3, but still are not globally good solutions. By contrast, the configuration on the right shows the kind of pattern reached using simulated annealing. The situations on the left panels of Figs. 3 and 5 are analogous (both showing systems that are stuck in local minima) as are the patterns on the right panels (both showing globally optimal solutions).

7.3. Manipulating superficial similarity

Superficial similarity was manipulated in a manner similar to Gentner and Toupin (1986). In their study, children were better able to retell stories when the characters involved in the retelling looked similar and played the same roles as those in the original story. In their "cross-mapping" condition, similar characters were used for the original story and its retelling, but the roles assigned to these characters were exchanged (see Ross, 1987 for another use of this technique). Likewise, our dissimilar

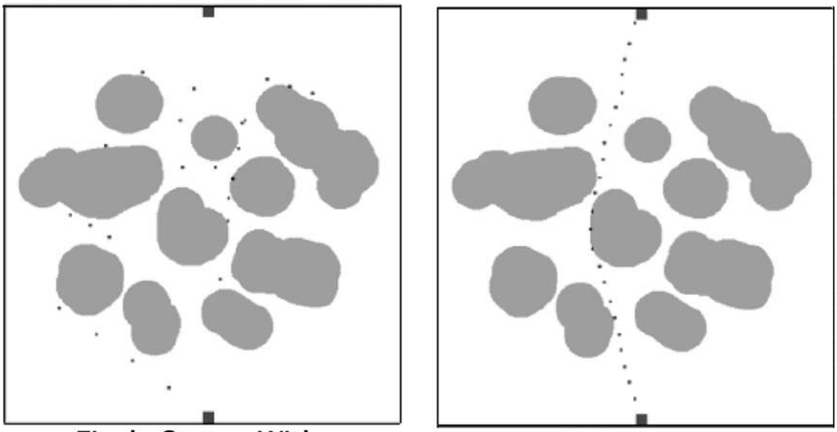

Fig. 5. Two possible final configurations for the "Path finder" simulation. If the balls move toward their neighbors without any randomness, then they will typically create "knots" that fail to form a single coherent pathway between the endpoints. If the amount of randomness is gradually reduced, then coherent pathways are formed. These two possibilities are analogous to the two possibilities shown in Fig. 3.

condition created a cross-mapping between balls and obstacles by swapping the colors of these objects across simulations. In the superficially similar condition, the colors for elements that played the same role were identical. By manipulating superficial similarity in this manner, the overall similarity between the simulations remains constant across similarity conditions; what changes is the level of similarity between elements that play analogous roles.

In addition to measuring the apprehension of the simulated annealing principle by multiple-choice quiz performance and strategic interactions with the simulations, participants were also tested on their ability to see the analogy between the two simulations. Previous research has shown that being able to abstractly express a schema that generalizes two problems is correlated with positive transfer from one problem to the other (Catrambone & Holyoak, 1989; Gick & Holyoak, 1983; Holyoak & Koh, 1987; Novick & Holyoak, 1991). In accord with this, one might predict that the performance improvement on the second simulation's quiz that is due to the first simulation might be correlated with a participant's ability to draw an explicit analogy between the simulations.

7.4. Method

7.4.1. Participants

Thirty-eight undergraduate students from Indiana University served as participants in order to fulfill a course requirement. The students were split evenly into the similar and dissimilar simulations conditions.

7.4.2. Materials

Participants were allowed to freely explore the simulations shown in Figs. 2–5. When presented with a simulation, participants were also given a one-page instruction sheet showing the rules by which the simulation operated, and a general goal or line of inquiry to pursue while exploring the simulation. For "Dropping balls," the goal was to find a way of having all of the balls settle on the lowest spot of a landscape without individually moving them. For "Path finder," the goal was to have the balls form a pathway connecting the fixed blue points and avoiding the green obstacles. Participants were given graphic examples of the goals. Participants were told that there were no hidden rules that described the behavior of the systems; everything that the balls did was governed by the given rules. Finally, participants were given descriptions of the important parameters of each simulation that were under their control.

Each simulation had a set of buttons and slider parameters on the left side of a 43-cm screen, and a graphics window on the right side of the screen. Participants could draw and erase obstacles and landscapes in the graphics window. By moving the mouse and pressing buttons, participants could directly affect the graphics window by drawing, erasing, moving, and placing agents. Buttons were used to reset the simulations, clear the screen, obtain help on using the simulations, turn on/off the simulations, and initiate the quiz. User-controlled sliders were used to control the continuously varying parameters of the simulations, such as the amount of randomness with which balls moved and the amount of movement on each time step. See Figs. 2–5 to see the parameters used in each of the simulations and their arrangement. Each simulation also contained a continuously updating graph that plotted how an important measure of performance in the simulation varied over time. Free versions of the Macintosh software can be downloaded at http://cognitrn.psych.indiana.edu/rgoldsto/complex/.

Each of the simulations had dynamically changing displays inside the graphics window. These displays were updated every 17 ms, and were instantaneously affected by user-controlled changes to parameter values. For example, as a participant reduced the randomness of ball movement in Dropping balls or Path finder, she or he would immediately see the balls move around with reduced jitter.

The simulations were all run on Power Macintosh G3 computers. The participants were run in groups of 8–12 in a large room containing 30 computers. The participants were separated so that they would distract each other as little as possible.

In the Path finder simulation the balls were red and the obstacles were green. For participants in the similar simulations condition, the balls in Dropping balls were red and the landscape was green. For participants in the dissimilar simulations condition, the balls in Dropping balls were green and the landscape was red.

7.4.3. Procedure

Participants were told that they would be exploring two computer simulations, but were not told that they were related in any way. For each of the simulations, they were first given an instruction page orienting them to the rules, parameters, and goal of the simulation. Then they were allowed to freely explore the simulation for 20 min.

During this time, three research assistants circulated around the room to answer questions that the students had about the rules governing the agents' behavior and the interface controls. In addition, the research assistants encouraged the participants to explore the simulations by manipulating parameters and by keeping the goal in mind. If a participant was clearly engaged in a digressive activity such as drawing their initials on the screen, they were encouraged to try to understand the behavior of the agents. During the exploration period, all key presses, mouse movements, and parameter changes were recorded by the computer, with time and duration information attached.

After the exploration period, participants were told to take the quiz by pressing the "Quiz" button. The computer then presented seven multiple choice quiz questions, samples of which are shown in Appendix A. Participants indicated their choice by clicking the mouse button while the cursor was inside a circle associated with the choice. After a choice was made, the computer automatically proceeded to the next question without presenting any feedback on the correctness of a response. Participants were not allowed to go back to earlier questions.

Participants were first given one of the two Dropping balls simulations, with either unchanged or swapped colors. After exploring Dropping balls and taking its associated quiz, participants were all transferred to the Path finder simulation with red balls and green obstacles. After completing both simulations, participants were given a multiple choice quiz probing their insight into analogous relations between the two simulations they explored. A definition of "analogy" was given, and an example was presented of the analogous relations between a solar system and an atom. Appendix B shows four of the six questions that were presented. Participants were shown each question on the computer screen, and responses were made by pressing the mouse button inside one of four circles displayed next to the four choices.

7.4.4. Results

Participants in the similar and dissimilar simulations conditions did not differ in their quiz performance on the first simulation (Dropping balls), producing respective accuracies of 40.3 and 40.6%, unpaired $t(36) = 0.23$, $p > .1$. However, with respect to performance on the second simulation, participants performed better in the dissimilar than similar condition, with respective accuracies of 59.1 and 40.8%, unpaired $t(36) = 2.8$, $p < .01$. On the quiz testing knowledge of analogous elements across the simulations, there was a non-significant trend for better performance in the dissimilar than similar condition, with respective accuracies of 46.1 and 39.3%, unpaired $t(36) = 1.8$, $p = .08$.

A secondary analysis was conducted to determine whether there were individual differences related to understanding that predicted the amount of transfer in the similar and dissimilar conditions. Based on quiz performances on the first simulation, participants from each similarity condition were divided into two groups based on whether a participant got fewer than three quiz answers correct (poor performers) or more than two answers correct (good performers). Eighteen and 20 participants were classified as good and poor performers, respectively. The number of participants in good-similar, good-dissimilar, poor-similar, and poor-dissimilar conditions were 10, 8, 9, and 11. The poor and good performers had mean quiz performances of 26.2 and 53.4% on the first simulation, respectively. There was an interaction between similarity and first simulation performance with performance on the transfer simulation as the dependent measure, as measured by a 2 (similar versus dissimilar) $\times$ 2 (good versus poor performance) factorial ANOVA (SPSS Univariate general linear model), $F(1, 34) = 7.8$, $p < .01$. As shown in Fig. 6, there was little difference between similar and dissimilar conditions for participants who performed well on the first quiz, but for poor performers, the dissimilar condition resulted in far better transfer performance than did the similar condition. There was not a significant interaction between similarity and first quiz performance on performance in the analogy quiz, $F(1, 34) = 1.8$, $p > .1$.

In addition to measuring quiz performance, we also analyzed participants' interactive usage of the simulations. All slider adjustments and button clicks were

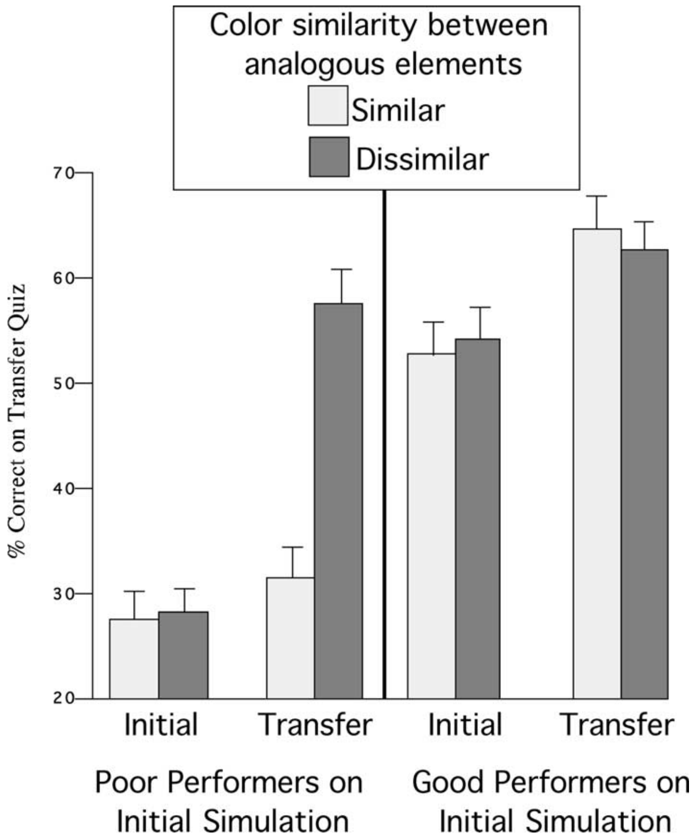

Fig. 6. Results from Experiment 1. Participants showed better quiz performance on the second, transfer simulation when the similarity of corresponding elements across the two simulations was reduced. A further breakdown of these results showed that this effect was completely due to participants who performed relatively poorly on the quiz for the original simulation.

recorded and time stamped, and this information was analyzed with respect to simulated annealing strategies. We defined a simulated annealing strategy as any situation where a participant set successively lower values to the randomness parameter. We defined the length of a simulated annealing strategy as the number of sequential parameter adjustments in which randomness values were progressively lowered. A long series of sequentially reduced randomness values is unlikely to be spontaneously produced by participants unless they were actively pursuing a simulated annealing strategy. Initial inspection of the results suggested that the most sensitive measure was the number of simulated annealing response patterns with length greater than 2, and in subsequent analyses this measure will be referred to as "Frequency of simulated annealing response strategy." For this measure, all key presses not directed toward adjusting the balls randomness were ignored.

Using this method of analyzing key presses, we measured the frequency of using a simulated annealing strategy. The frequency of using a simulated annealing response strategy was correlated with quiz performance for both the initial simulation (Pearson's $r = .29$, $p < .01$) and the transfer simulation ($r = .23$, $p < .01$). As shown in Fig. 7, good performers on the initial quiz use a greater number of simulated annealing response strategies than poor performers, on both the initial simulation, unpaired $t(36) = 3.7$, $p < .01$, and the transfer simulation, unpaired $t(36) = 3.2$, $p < .01$. There was a greater number of simulated annealing response strategies for the transfer than initial simulation, paired $t(37) = 2.5$, $p < .05$. There was not a significant interaction between similarity and first simulation quiz performance with number of simulated annealing responses on the second simulation as the dependent variable, $F(1, 34) = 2.1$, $p > .1$.

7.4.5. Discussion

For no group of participants was transfer better between relatively similar simulations, and for the poor initial performers, transfer was significantly better for relatively dissimilar simulations. This relatively good transfer occurs between conditions that Gentner and Toupin (1986) would call a "cross-mapping." In the dissimilar condition, simulations linked by a common abstract principle had surface-level features that would suggest an improper mapping between the elements of the two simulations. In this condition, salient color information would tend to de-emphasize the correspondence between the balls in the simulations, and would suggest a connection between the balls in one simulation and the obstacles in the other simulation. Previous research has shown a debilitating effect of cross-mapped surface information (Gentner & Toupin, 1986; Ross, 1987). In the current experiment, reducing the surface similarity between analogous elements across simulations produced no worse appreciation of the analogical correspondences and produced better performance on the second simulation for the initially poor performers.

One important difference between several previous experiments that involve cross-mappings (e.g., Ross, 1987) and Experiment 1 is that Experiment 1's cross-mapped feature, color, is not likely to be intrinsically connected to objects. Obstacles and balls can be either green or red. In Ross (1987), the cross-mapped items, such as cars and mechanics, are semantically rich and their meaning affects expectancies about the role

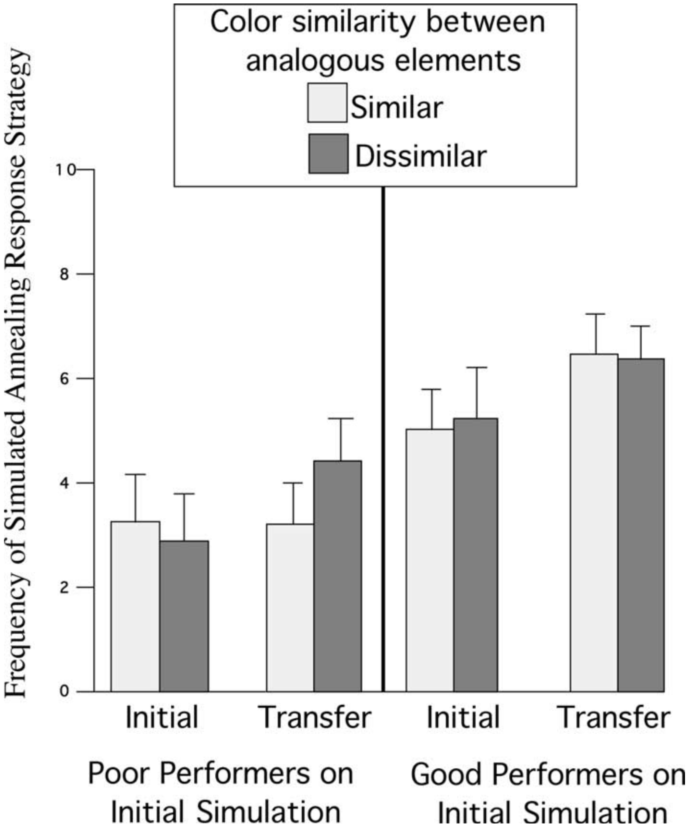

Fig. 7. Simulated annealing response strategies were measured during learners' interactions with the simulations. This graph shows the interaction between a learner's initial quiz performance level and the similarity between simulations, using these response strategies as a measure of performance.

they will play in their equations (Bassok, 1996; Bassok, Wu, & Olseth, 1995). One reason why we did not find the deleterious effects of cross-mapping found by others (Gentner & Toupin, 1986; Ross, 1987) may be that our cross-mappings did not go against semantic expectancies or promote incompatible interpretations of objects.

One account for the better transfer in the dissimilar, relative to similar, condition for poor performers is that these participants, when in the similar condition, derived a relatively superficial understanding of the connection between the two simulations precisely because of their high level of surface similarity. For these participants, the two simulations would have matched very well at a surface level, and participants may have remained at this level of understanding because of its apparent success in relating the simulations. Similar to Goldstone et al.'s (1991) participants increasing their attention to surface over abstract commonalities when additional surface commonalities were added, a high level of surface similarity may have promoted a relatively surface-level understanding of the simulations. This would have impaired quiz performance because the quiz questions were designed to test participants' general appreciation of simulated annealing. Another way of expressing this idea is that the deep, abstract commonalities between situations may be concealed if the situations also share many salient surface-level commonalities (Gick & Holyoak, 1983).

This account must be supplemented with an account that explains why participants' performance on the first simulation interacted with simulation similarity. One of the predictions described in Section 5 was that poor performers would be most likely to have a superficial understanding of the simulations, and that their performance on the transfer simulation would be best if surface-level properties were preserved across simulations. In contrast, good performers are presumably less tied to surface representations (Chi et al., 1981; Miller & Stigler, 1991; Schiano et al., 1989) and would be better able to see abstract commonalities across superficially dissimilar simulations than poor performers.

In fact, our results showed a very different kind of interaction, in which good performers showed little effect of similarity (consistent with the above prediction), but poor performers were much better with dissimilar rather than similar simulations. If poor performers have difficulty ignoring surface similarities when they should be focusing on abstract commonalities, this difficulty should be most pronounced when the surface similarities are compelling. Poor performers may do poorly with similar simulations because this situation leads them to construe the second simulation at a superficial level. Good performers instead may be better able to construe a situation at an abstract level regardless of misleading superficial similarities. To summarize, our results indicate that poor performers are more influenced by superficial similarity than good performers. However, this greater influence is not because poor performers need supporting superficial similarities to see the connection between abstractly related situations, but rather because superficial similarities distract poor performers from identifying and using abstract similarities. The task of characterizing situations when superficial similarities support versus interfere with an abstract understanding will be reserved for Section 11.

8. Experiment 2

Experiment 1 manipulated the surface similarity between two simulations by either preserving or swapping the colors of corresponding elements across the simulations. This manipulation controls for the overall similarity between the conditions, selectively manipulating the surface similarity between elements that play analogous roles. However, this manipulation produces a cross-mapping situation that may be somewhat unnatural. It may not be very common for elements X and Y from Simulation 1 to be analogous to elements X' and Y' from Simulation 2, and yet for there to be a dimension along which X is more similar to Y' than X', and Y is more similar to X' than Y'. A more natural way of manipulating surface similarity is simply to add to or subtract from the surface similarities between two analogous elements (Gick & Holyoak, 1983; Ross, 1987). Experiment 2 is an attempt to replicate Experiment 1 using this method of manipulating surface similarity. In particular, we first allow students to explore Dropping balls, and then transfer students to Path finder. In Path finder, the agents potentially undergoing simulated annealing always appear as blue marbles. In the Similar condition, the agents undergoing simulated annealing in Dropping balls are also blue marbles. In the Dissimilar condition, they are black and white soccer balls. The obstacles in both simulations are green, solid patches. Thus, both the overall similarity between the simulations and the similarity between one pair of corresponding elements are increased from the Dissimilar to Similar condition.

8.1. Method

8.1.1. Participants

Forty-four undergraduate students from Indiana University served as participants in order to fulfill a course requirement. The students were split evenly into the Similar and Dissimilar conditions.

8.1.2. Materials

The simulations were the same as those used in Experiment 1, except for the superficial appearance of the agents. Whereas Experiment 1 used small dots to designate the agents in both simulations, Experiment 2 used 0.9-cm drawings. As illustrated in Fig. 8, agents were always depicted by blue marbles during Path finder. In the Similar condition, the same blue marbles were used as agents during Dropping balls. In the Dissimilar condition, the blue marbles were replaced by black and white soccer balls during Dropping balls.

8.1.3. Procedure

The procedures were the same as those used in Experiment 1. Participants were always exposed to Dropping balls, then quizzed on Dropping balls, then exposed to Path finder, then quizzed on Path finder, and then quizzed on their understanding of the analogical relations between the two simulations.

8.1.4. Results

Participants in the similar and dissimilar simulations conditions did not differ in their quiz performance on the first simulation (Dropping balls), producing respective accuracies of 40.0 and 38.7%, unpaired $t(42) = 0.14$, $p > .1$. However, participants

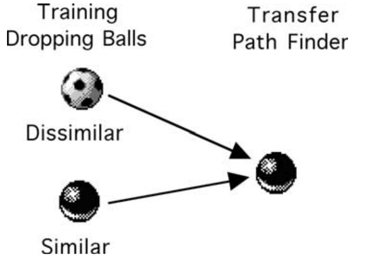

Fig. 8. The design for Experiment 2. The balls in the initial "Balls Dropping" simulation were either similar or dissimilar to the analogous balls of the transfer simulation "Path finder."

performed better in the dissimilar than similar condition on the transfer simulation (Path finder), with respective accuracies of 51.1 and 41.8%, unpaired $t(42) = 2.9$, $p < .01$. On the quiz testing knowledge of analogous correspondences across the simulations, similar and dissimilar conditions yielded accuracies of 45.8 and 41.2%, respectively, unpaired $t(42) = 1.4$, $p > .1$.

As with Experiment 1, a more refined analysis of participants' performance was obtained by dividing participants into those who performed well and poorly on the initial simulation quiz. Based on quiz performance on Dropping balls, participants from each similarity condition were divided into two groups based on whether a participant got fewer than three quiz answers correct (poor performers) or more than two answers correct (good performers). The number of participants in good-similar, good-dissimilar, poor-similar, and poor-dissimilar conditions were 12, 13, 11, and 10. The poor and good performers had mean quiz performances of 25.9 and 50.7%, respectively, on the initial simulation. There was an interaction between similarity and first simulation performance with performance on the second, transfer simulation as the dependent measure, $F(1, 40) = 5.8$, $p < .05$. As shown in Fig. 9, there was little difference between similar and dissimilar conditions for participants who performed well on the first quiz, but for poor performers, the dissimilar condition resulted in far better transfer performance than did the similar condition. This pattern is very similar to that obtained in Experiment 1, although the magnitude of the interaction was somewhat attenuated. There was not a significant interaction between similarity and first quiz performance on performance in the analogy quiz, $F(1, 40) = 1.8$, $p > .1$.

The frequency of using a simulated annealing response strategy was measured by the number of situations in which a participant successively reduced randomness two or more times in a row. This frequency was correlated with quiz performance for both the initial simulation (Pearson's $r = .33$, $p < .01$) and the transfer simulation ($r = .30$, $p < .01$). As shown in Fig. 10, good performers on the initial quiz use a greater number of simulated annealing response strategies than poor performers, on both the initial simulation, unpaired $t(42) = 4.1$, $p < .01$, and the transfer simulation, unpaired $t(42) = 3.5$, $p < .01$. There was a greater number of simulated annealing response strategies for the transfer than initial simulation, paired $t(43) = 2.9$, $p < .01$. Unlike Experiment 1, there was a significant interaction between similarity and first simulation quiz performance with number of simulated annealing responses on the transfer simulation as the dependent variable, $F(1, 40) = 5.0$, $p < .05$. The interpretation of this interaction is the same as the interpretation for the interaction involving transfer quiz performance. In particular, poor performers are more likely to use a simulated annealing strategy for the transfer simulation when this simulation is dissimilar to the initial simulation rather than similar. This trend is not found for good performers.

8.1.5. Discussion

The results from Experiment 2 generally replicate those obtained in Experiment 1, using a different method for manipulating the similarity between initial and transfer simulations. In particular, both experiments found better transfer between dissimilar

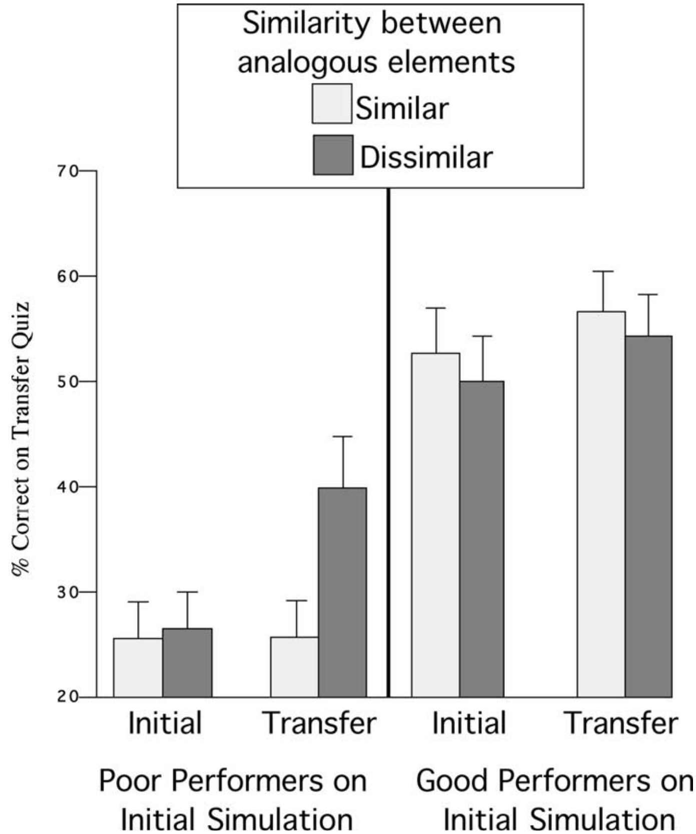

Fig. 9. Results from Experiment 2. Participants who performed relatively poorly on the initial simulation's quiz showed better transfer to an analogous simulation when the overall similarity between the simulations was relatively low.

than similar simulations, but also found that this main effect was modulated by an interaction involving the participants' performance on the first simulation. The form of this interaction was also similar to Experiment 1, with a larger difference between similar and dissimilar transfer conditions for relatively poor performers on the first simulation than for good performers. This interaction is surprising in that good performers might be expected to be less susceptible to disruptions in superficial similarity than poor performers if poor performers rely on these similarities to the exclusion of more abstract similarities (e.g., Chi et al., 1981). However, this account has difficulty with even the main effect that reducing superficial similarity seems to have a facilitative, not disruptive, effect on abstraction-based transfer.

As argued for Experiment 1, an alternative account is that superficial similarities have a disruptive influence because they drive participants to create inappropriately surface-based representations for the transfer simulation. These surface-based representations compete against the more abstract representations that are required to correctly answer the quiz questions. Poor performers are relatively prone to interpret

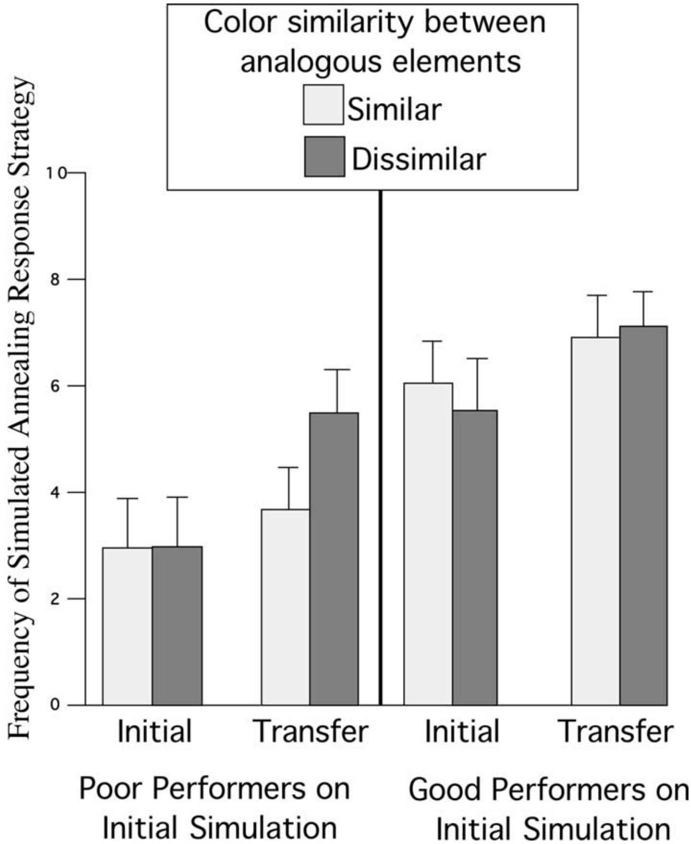

Fig. 10. The results from learners' response strategies, diagnosed by their sequence of key presses, supported the quiz results. Poor performers used more simulated annealing response strategies when the transfer simulation was dissimilar rather than similar to the initial simulation.

simulations superficially, and thus are at highest risk for being swayed by the superficial similarities in the Similar condition. Good performers are better able to focus on abstract properties, ignoring superficial similarities. Thus, this argument uses the same premise as Chi et al. (1981) that poor performers are strongly influenced by superficial features, but applies this premise to a situation where superficial similarities are hindrances rather than aids to abstraction-based transfer.

9. Experiment 3

The observed amount of transfer between simulations may depend on how performance is measured. Explict and implicit measures of performances can be distinguished, where explicit measures of performance require participants to verbally express a solution or understanding and implicit measures only require that participants behaviorally display understanding without necessarily being able to put this understanding into words. Thus far, the experiments have measured transfer between simulations by gauging participants' quiz performance on the transfer simulation, and by examining the strategic interactions with the transfer simulation as evidenced by key presses. Quiz performance is a fairly explicit measure of transfer in that it requires participants to select a verbal answer from a set of alternatives. A more implicit measure of transfer is to observe whether participants can better solve problems in the transfer simulation when it has been preceded by a related simulation (Schunn & Dunbar, 1996). Participants who have a hard time expressing in words the simulated annealing principle may still be able to apply knowledge gained from the initial simulation to improve their ability to solve posed problems during the transfer simulation. The primary purpose of Experiment 3 is to explore transfer between complex systems simulations using this implicit measure of transfer. In particular, when transferred to the second simulation, participants are given problems to solve that require a simulated annealing strategy. Performance is measured by the time required to supply solutions to these problems.

There is already strong evidence for performance-based transfer between situations in the absence of explicit appreciation of the connection between them (Underwood, 1996). For example, transfer between visual search tasks based on complex rules is observed even when participants show no awareness of the rules (Stadler, 1989). However, it is less clear how the explicitness of a measure of transfer interacts with the superficial similarity between the simulations. Promoting superficial similarity between simulations might be predicted to promote transfer based on implicit measures more than explicit measures. Implicit measures of learning often times show a greater dependence on the superficial similarity between training and testing contexts (see Schacter, 1987 for a review). If so, then the disadvantageous effects of high superficial similarity found in Experiments 1 and 2 may not be found for an implicit measure of performance.

Another minor change from Experiment 2 to Experiment 3 is that we ask our learners to give their Scholastic Achievement Test (S.A.T.) scores. In assessing the difference between poor and good comprehenders, it is useful to know whether this difference is specific to the simulations, or is an indication of a more general individual difference. Poor comprehenders of a simulation might be expected to perform relatively poorly on other measures of academic achievement. If it is the case that poor comprehension is correlated with generally poor achievement, then statistical efforts to disentangle the two kinds of individual differences can reveal which of them is more closely tied to the observed advantage for dissimilar as compared to similar simulations.

9.1. Method

9.1.1. Participants

Ninety undergraduate students from Indiana University served as participants in order to fulfill a course requirement. The students were split evenly into the Similar and Dissimilar conditions.

9.1.2. Procedure

The simulations and materials followed those used in Experiment 2, with the exceptions described here. Participants were first asked to report their cumulative S.A.T. or Academic College Test (A.C.T) scores. Cumulative, rather than itemized scores were requested because a polling of participants suggested much better memory for cumulative scores.

All participants were first given 20 min experience with a version of Dropping balls in which the balls either appeared as soccer balls (Dissimilar condition) or blue marbles (Similar condition). During this initial simulation, participants were encouraged to explore the simulation and were also given the general goal of developing a method for having all balls fall to the lowest region of a drawn landscape. After the previously used multiple-choice quiz on this simulation, participants were all transferred to a version of Path finder in which the balls appeared as blue marbles.

During the Path finder simulation, participants were given instructions on how to use the simulation, and a description of the general goal of finding short pathways around obstacles to connect endpoints. Several functions of the simulation were disabled relative to previous experiments. The only sliders that participants could manipulate were those that controlled the balls' randomness and amount of movement. The only accessible buttons were those for resetting and randomizing the balls' positions, and for turning on/off the adjustment of balls. Participants were not allowed to draw or erase obstacles. These functions were disabled so that participants would not be able to solve presented problems by changing the problems rather than by using a simulated annealing strategy.

Participants were presented with three problems to solve, in increasing order of difficulty. In the first problem, there was a circular obstacle in the middle of the field covering 8% of the field. In the second problem, there was an oval obstacle centered in the right-middle portion of the field, covering 14% of the field. In the third problem, three ovals occupied 17% of the field and were arranged so that the optimal path required three bends.

By adjusting the parameters associated with the amount of randomness and movement of the balls, participants could solve each of the three problems. The ideal solution method for each of three problems was the same. The amount of movement should be kept relatively large. The amount of randomness should be initially high and should be gradually reduced. The computer automatically detected when a successful solution to a problem was reached, and advanced students to the subsequent problem. A successful solution was defined as one in which the average distance between balls was no more than 5% greater than optimal, and in which no pair of neighboring balls was separated by more than three times the average distance. After 20 min or the successful completion of all three problems, participants were given the 7-item multiple choice quiz used in Experiment 2.

Following both simulations and quizzes, participants were asked whether they spontaneously noticed similarities between the two simulations. Specifically, they were asked, "As you were exploring the Path finder simulation, were you spontaneously reminded of the Dropping balls simulation? Check off any of the similarities that you noticed." Participants then were given the following four choices and were asked to select as many as relevant: "The appearance of buttons, sliders, and graphics," "The way that I used the simulation to draw objects, move balls around, clear the screen, etc.," "The importance of having balls move with randomness," and "The strategy of finding good solutions by gradually reducing the randomness in the balls' movements."

9.1.3. Results

Participants in the similar and dissimilar simulations conditions did not differ in their quiz performance on the first simulation (Dropping balls), producing respective accuracies of 42.8 and 40.5%, unpaired $t(88) = 0.23, p > .1$. Participants also did not perform significantly different in the dissimilar than similar condition on the transfer quiz (Path finder), with respective accuracies of 51.7 and 49.4%, unpaired $t(88) = 1.1, p > .1$.

Based on quiz performance on Dropping balls, participants from each similarity condition were divided into two groups based on whether a participant got fewer than three quiz answers correct (poor performers) or more than two answers correct (good performers). Forty-eight and 42 participants were classified as good and poor performers, respectively. The number of participants in good-similar, good-dissimilar, poor-similar, and poor-dissimilar conditions were 25, 23, 20, and 22. The poor and good performers had mean quiz performances of 26.8 and 54.5%, respectively, on the initial simulation. There was an interaction between similarity and first simulation quiz performance with performance on the second, transfer quiz as the dependent measure, $F(1, 86) = 4.9$, $p < .05$. As shown in Fig. 11, poor performers on the first quiz performed better in the dissimilar than similar condition, but the opposite trend was found for good performers. Relative to Experiment 2, performance was somewhat better for the dissimilar rather than similar condition, but the magnitude of the interaction was comparable.

A second way of measuring performance on the transfer simulation is by analyzing the time required to solve the three posed problems. Eighty-two out of the 90 participants managed to solve all three problems within the allotted 20-min period. For the remaining eight participants, the maximum time of 20 min was used as the estimate of their solution time. Fig. 11 shows the solution times required for poor and good initial quiz performers, and for participants in the similar and dissimilar conditions. There was a significant interaction between similarity and first simulation quiz performance with time to solve transfer problems as the dependent variable, $F(1, 86) = 5.4$, $p < .01$. This interaction is consistent with the interaction found for quiz performance. Poor initial quiz performers require a particularly long time to solve the transfer simulations' problems when the two simulations were similar, rather than dissimilar, in their superficial appearances.

Out of the 90 participants, 82 reported taking the S.A.T. and eight reported taking the A.C.T. The average scores for the A.C.T. and S.A.T. were 26.4 and 1108, respectively. The scores from these two achievement tests were normalized by dividing by the total number of points possible for each test (36 for the A.C.T and 1600 for the S.A.T.). As shown in Table 1, Pearson correlations were calculated between every pair of the following variables: time required to solve transfer problems, normal-

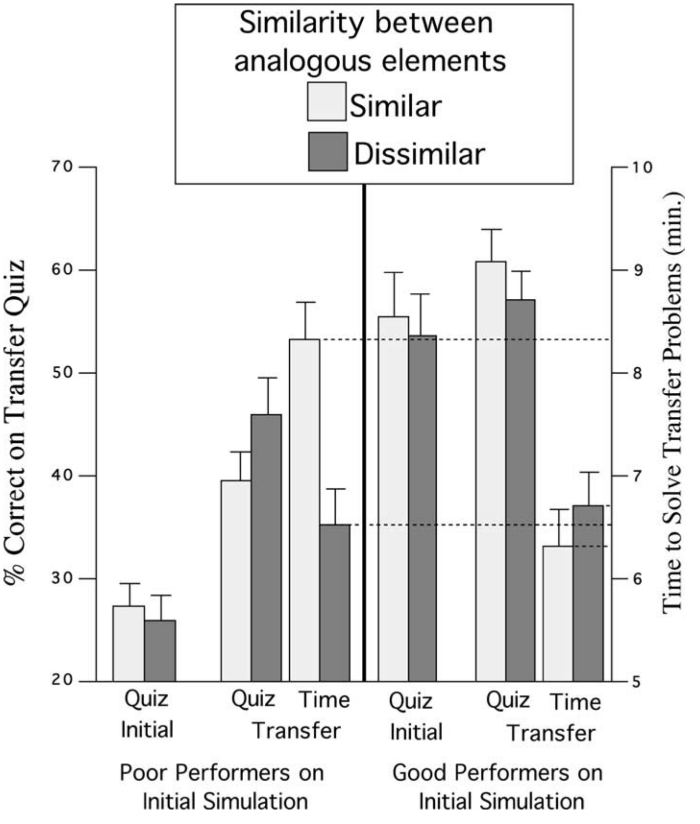

Fig. 11. Quiz performance and response strategy results from Experiment 3.

Table 1
Correlations and significance levels from Experiment 3

	Time required to solve transfer problems	Achievement test score	Initial quiz performance	Final quiz performance
Time required to solve transfer problems		$r = -.07$ $p = .781$	$r = -.15$ $p = .092$	$r = -.26$ $p = .027$
Achievement test score			$r = .301$ $p = .014$	$r = .240$ $p = .032$
Initial quiz performance				$r = .384$ $p = .002$

ized achievement test score, initial quiz score, and final quiz score. These correlations reveal that achievement test scores are significantly correlated with both quiz scores but not with the time to solve transfer problems. The highest correlation occurred between initial and final quiz performances. Time to solve transfer problems was significantly correlated with final quiz performance, and this correlation continued to be significant even when the variability in final quiz performance due to achievement score was partialled out ($p < .05$). When asked at the end of the experiment, learners often claimed to have spontaneously noticed similarities between the simulations. The average percentage of learners affirming that they noticed similarities related to graphics, methods of interaction, randomness, and simulated annealing strategy were 93, 86, 37, and 24%, respectively. These percentages did not vary significantly as a function of simulation similarity, unpaired $t(88) = 1.5$, $p > .1$. The noticed similarities were grouped into relatively shallow (the first two questions) versus relatively deep (the last two questions) remindings. There was an interaction between type of reminding and initial quiz performance level, with good performers more likely to report deep remindings than poor performers, $F(1, 86) = 5.2$, $p < .01$. The percentages of good-shallow, good-deep, poor-shallow, and poor-deep remindings were 90, 37, 89, and 29%, respectively.

9.1.4. Discussion

Experiment 3 largely replicated the interaction between participants' performance and simulation similarity observed in Experiment 2. In terms of their performance on the transfer simulation, poor performers on the first quiz were hurt, not helped, by increasing the superficial similarity between simulations. This effect was, if anything, more robust for the problem-based measure of performance than it was for the quiz-based measure primarily used in Experiments 1 and 2. The earlier results generalize to implicitly measured transfer between abstractly related simulations.

There is a significant correlation between the implicit, performance-based measure of transfer, and the quiz-based measure of transfer. This provides grounds for believing that the previous use of quiz-based transfer at least partially taps into learners' ability to apply abstract principles from one domain to another. On the one hand, these two measures continue to be correlated with one another even after the influence of high school achievement test score has been factored out. On the other hand, the correlations are not absolutely very high, and the strongest correlation is between initial and final quiz performances. This latter correlation may be high because both quizzes measure students' expressible, rather than implicit, knowledge. The questions asked in the initial and final quizzes were designed to be analogous, and this is likely to have further strengthened the correlation between the two quiz performances.

The results of the reminding questionnaire at the end of the experiment suggest that a high percentage of learners were reminded of the first simulation as they interacted with the second simulation. As discussed further in Section 11, this is consistent with our account of when superficial similarities are likely to interfere rather than promote abstract transfer. Superficial similarities are expected to be useful in situations where an earlier relevant situation may not be brought to mind as relevant to participants when solving a problem (Gick & Holyoak, 1980; Ross, 1984). For our experiment, the spontaneous rate of reminding was high, and the benefit of high superficial similarity in promoting reminding may not have been very advantageous in this context.

10. Experiment 4

Experiments 1–3 explored the role of superficial similarity in promoting or interfering with the transfer of abstract principles between analogous simulations. Another method of investigating the supportive and competitive relations between superficial and abstract representations is by manipulating the amount and richness of superficial information (Ratterman & Gentner, 1998). An assumption underlying much of the development of computer simulations is that information is more effectively understood when it is presented in a concrete and vivid manner. As reviewed in Section 1, human reasoning is often better with concrete rather than abstract materials. However, concreteness may also interfere with the development of abstract, context-independent representations (Anderson et al., 1996). Increasing the concreteness of an object sometimes makes it less likely to be used as an analogical device for understanding a relationally similar situation (DeLoache, 1991, 1995). DeLoache has described these empirical results as suggesting a competition between viewing an object as a symbol for representing another object and as a concrete object in its own right. Increasing the richness of superficial commonalities between scenarios leads children away from appreciating deeper commonalities if the superficial and deep commonalities do not consistently suggest the same correspondences between scenarios (Ratterman & Gentner, 1998). In addition, more abstracted, idealized word problems can yield better transfer to concrete problems than vice versa (Bassok & Holyoak, 1989). Experiment 4 investigated whether a relatively abstract simulation resulted in better abstract quiz performance on the simulation itself compared with a more concrete simulation, and whether this training transferred better to an analogous simulation.

The scientific principle explored in Experiment 4, competitive specialization, is different from the one used in Experiments 1–3. Parts that start out homogeneous and undifferentiated can each become specialized, as a result of interactions between the parts (O'Reilly, 2001, von der Malsburg, 1973). A general application of self-organization is the allocation of resources to cover an abstract or concrete territory. It is often optimal for different agents to be specialized for different regions. In these situations, a good solution is found if every region has an agent reasonably close to it (Stephens & Krebs, 1987).

An elegant solution to problems of (close to) optimal covering such as these is to allocate agents (e.g., neurons, oil drills, and flies) by executing the following three steps repeatedly: (1) randomly selecting a resource from among the entire set of resources to be covered, (2) determining the closest agent to this resource, and (3) adapting this closest agent toward the resource at a relatively fast rate and adapting all other agents toward the resources at a relatively slow rate. This algorithm works by creating adapted and differentiated agents. If all agents adapted as quickly as the closest agent, then, ironically, they would cover the territory less optimally as a group, because all of the agents would occupy a similar position at the territory's center-of-mass.

10.1. Ants and food

The first example of competitive specialization involves ants foraging food resources drawn by a user. The ants follow exactly the three rules described above. At each time step, a piece of food is randomly selected, and the ant closest to the food moves with one rate, and all of the other ants move with another rate. Similar to the earlier simulations, a learner can reset the ants' positions, clear the screen of food, place new ants, move ants, start/stop the ants' movements, and set a number of simulation parameters. The two most critical user-controlled parameters determine the movement speed for the ant that is closest to the selected food (called "closest rate" in Fig. 12) and the movement speed for all other ants ("Not closest rate").

Starting with the initial configuration of three ants and three food piles shown in Fig. 12, several important types of final configuration are possible and are shown in Fig. 13. If only the closest ant moves toward a selected piece of food, then this ant will be the closest ant to every patch of food. This ant will continually move to new locations on every time step as different patches are sampled, but will tend to hover around the center of gravity of the food patches unless the adjustment rate is set very high. The other two ants will never move at all because they are never the closest ant to a food patch. This configuration is sub-optimal because the average distance between a food patch and the ant it is closest to (a quantity that is continually graphed) is not as small as it would be if each of the ants specialized for a different food pile. If

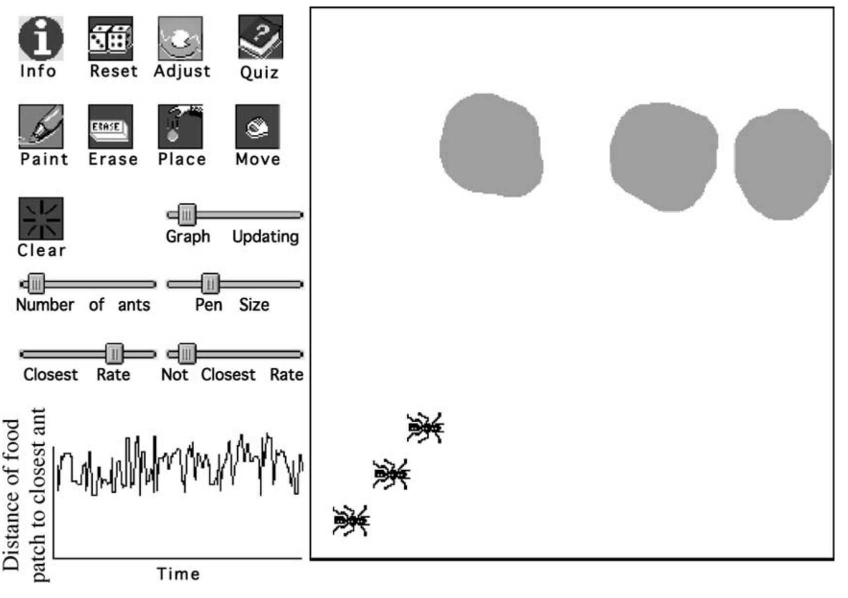

Fig. 12. A screen-dump of an initial configuration for "Ants and food." At each time step, a patch of food is randomly selected, and the ant closest to the patch moves toward the patch with one speed (specified by the slider "Closest Rate") and the other ants move toward the patch with another speed ("Not Closest Rate").

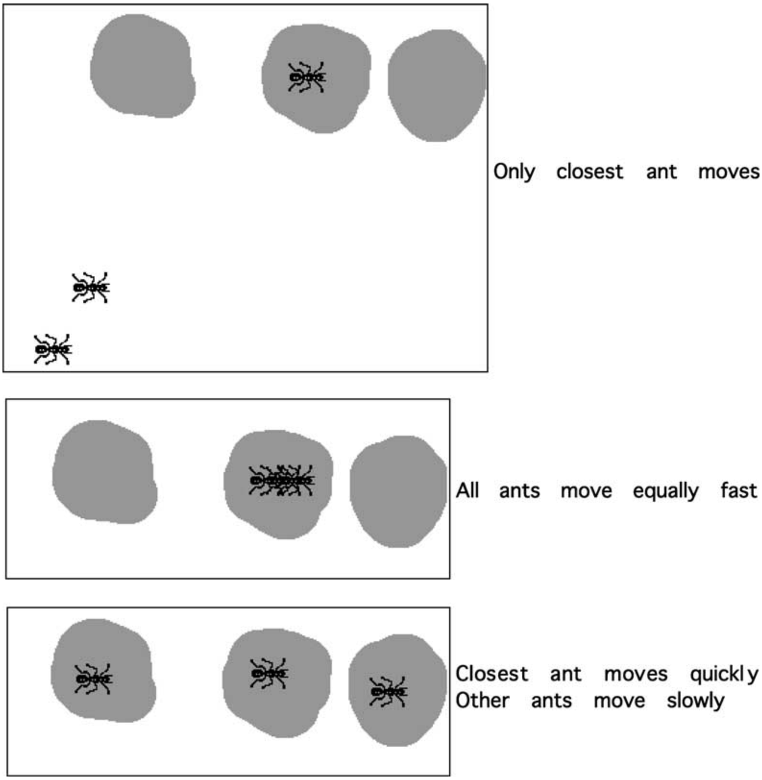

Fig. 13. If only the ant closest to a selected food patch moves, and if all of the patches are fairly close, then often a single ant will move toward the average position of all of the food patches. If all ants move equally quickly, then all ants will move toward the average position. If the ant closest to a selected food patch moves much faster than the other ants but all ants move a bit, then each of the ants will become specialized for one food patch. This third possibility illustrates competitive specialization.

all of the ants move equally quickly, then they will quickly converge to the same screen location. This also results in a sub-optimal solution because the ants do not cover the entire set of resources well. Finally, if the ant closest to a selected patch of food moves quicker than the other ants, but the other ants move too, then a nearly optimal configuration is achieved. Although one ant will initially move more quickly toward all selected food patches than the other ants, eventually this ant will move toward a patch of food, thereby distancing itself from another patch of food that will then be controlled by another ant.

10.2. Pattern learning

The second example of competitive specialization is more abstract, dealing with the development of categories of visually presented patterns. The simulation is based on the unsupervised neural network learning algorithm of competitive learning

(Rumelhart & Zipser, 1985). It is often desirable to have a system create categories that naturally capture the systematicities in a set of patterns. For example, if we present a system with examples of the letter "A" and examples of the letter "B," one useful thing for the system to do is to develop two categories—one for each type of letter. Once developed, these categories can be used as an efficient way of coding new inputs of the same type. One way of automatically creating appropriate categories is to randomly initialize the categories, and then repeatedly select a picture, find the category that is most similar to the selected picture, adjust this category so that it even more closely resembles the selected picture, and adjust all of the other categories at a slower rate.[1] This technique does not always produce optimal categories (for improvements to this algorithm, see Goldstone, in press), but usually results in categories that resemble the major categories implicit among the pictures. Learners interacting with the simulation can control the rates by which the most similar category (the "winner") and the other categories (the "losers") adjust toward a selected picture, draw and edit pictures, set the number of pictures and categories, and start/stop the adjustment of categories.

Fig. 14 shows an initial configuration of randomized categories, and three pictures of letters that were drawn by a user. Fig. 15 shows three potential category configurations resulting from different parameter settings. If the adaptation rate for the winning category is positive but the rate for losing categories is zero, then a single category will adapt toward all three letters (which are similar because of their common black backgrounds), leaving the other two categories unchanged. The single winning category will become a blend of all three letters, and consequently none of the letters will have a category that closely resembles it. The second panel shows the categories that emerge when all categories adapt equally quickly toward a presented picture. All categories will quickly become an identical blend of the same three pictures. The categories become progressively more similar to each other rather than differentiated over time because they are influenced by each presented picture in the same way. The third panel shows the category differentiation that occurs when the winning category adapts rapidly while the losing categories adapt much more slowly. Now, each category becomes specialized for one and only one of the pictures. Although one category initially adapts more quickly toward all three pictures than the other categories, eventually the other categories will be equally close to the pictures. At this point, a picture will attract a category, and this attraction will pull the category away from other pictures, leaving them open to be covered by a different category.

The analogy between this situation and the ants foraging for food is hopefully clear. The three panels of Fig. 13 are analogous to the respective panels in Fig. 15. The two domains are both instantiations of the principle of competitive specialization and are governed by the same mathematical formalism (Kohonen, 1995).

[1] Although it might seem that determining the closest agent to particular resource patch requires centralized leadership, Grossberg (1976) has shown how the closest agent can be identified on the basis of solely local interactions among the agents.

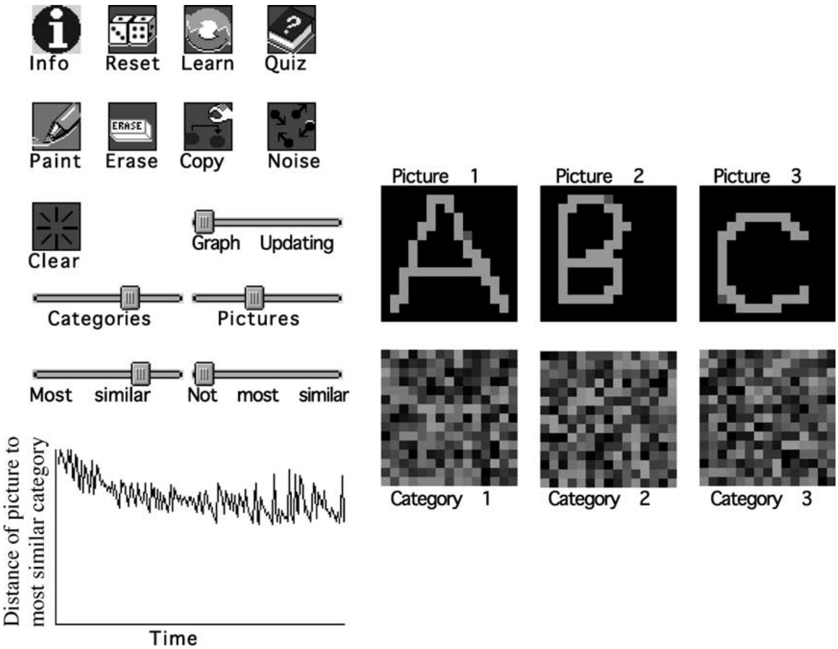

Fig. 14. A screen-dump for the simulation "Pattern learn." Users draw pictures, and prior to learning, a set of categories are given random appearances. During learning, a picture is selected at random, and the most similar category to the picture adapts its appearance toward the picture at one rate (specified by the slider "Most similar") while the other categories adapt toward the picture at another rate ("Not most similar").

10.3. Concreteness in competitive specialization simulation

Simulation concreteness was manipulated by changing the specificity of the graphics. Both concrete and idealized simulations involved the same dynamic, graphical displays used in the previous experiments. The only difference involved the appearance of the ants and food in the first simulation. In the concrete condition, a line drawing of a black ant and food identifiable as pieces of fruit depicted the agents and resource, respectively. In the idealized condition, ants were represented by black dots, and green patches represented food sources. In this manner, both simulations were equally graphical, but differed in the immediacy of the link between the graphical elements and objects they represented.

The results from Experiments 1–3 do not have a necessary implication for predictions in Experiment 4. However, if poor performers had difficulties in conditions with surface-level similarities across simulations in Experiments 1–3, then it might be expected that the concrete simulations would produce relatively poor abstraction-based transfer for them. That is, if the primary hurdle that participants must overcome in showing positive transfer is the salience of surface-level properties distracting them from abstract properties, then both high surface similarity and vivid surface properties would be expected to hinder abstraction-based transfer.

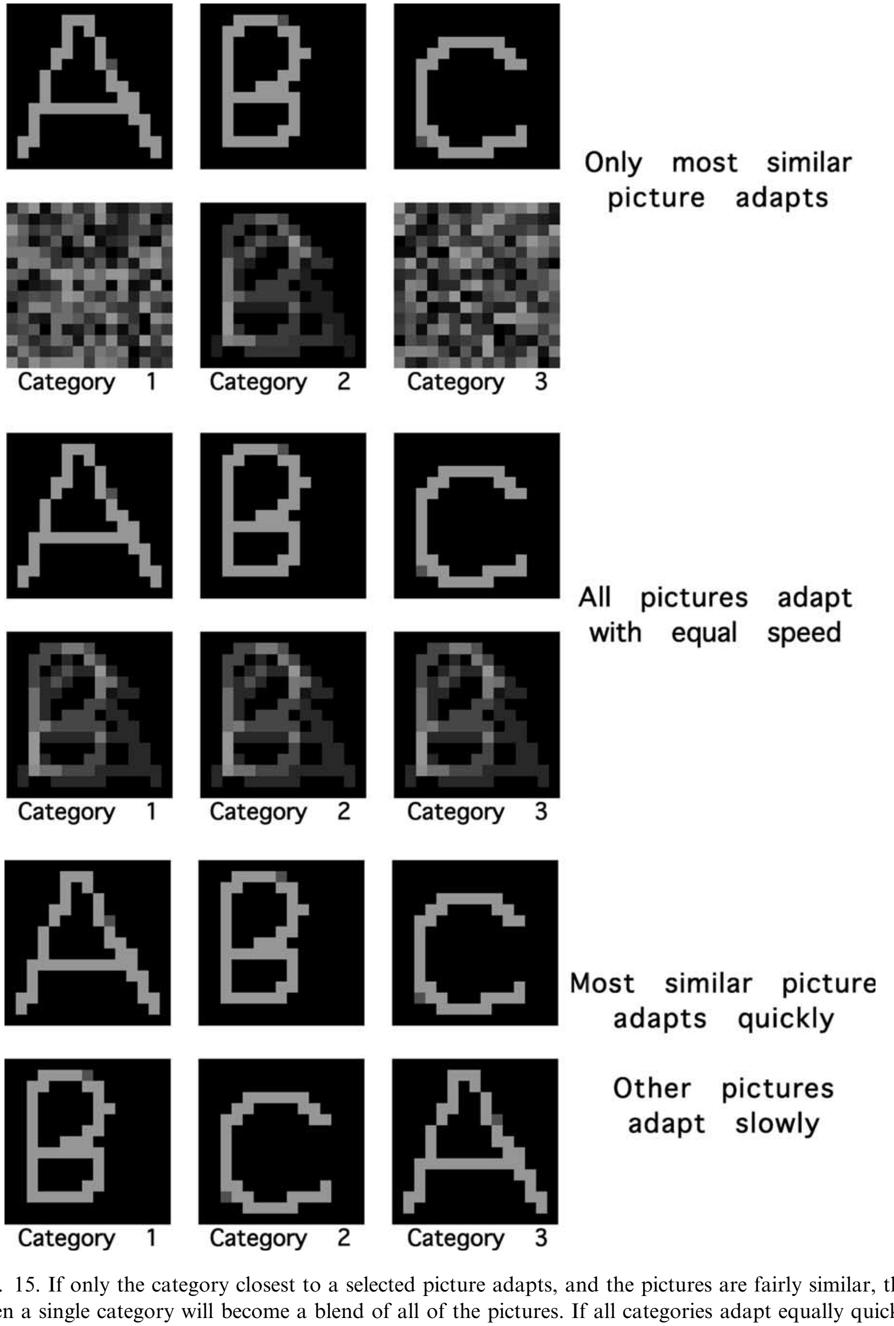

Fig. 15. If only the category closest to a selected picture adapts, and the pictures are fairly similar, then often a single category will become a blend of all of the pictures. If all categories adapt equally quickly, then each category will become a blend of all of the pictures. If the category most similar to a selected picture adapts much more quickly than the other categories but all categories adapt a bit, then each category will become specialized for one picture. These three outcomes are analogous to the three outcomes shown in Fig. 13.

10.4. Method

10.4.1. Participants

Sixty-one undergraduate students from Indiana University served as participants in order to fulfill a course requirement. Using a randomized assignment strategy, 29 students were randomly placed in the idealized graphics condition and the remaining students were placed in the concrete graphics condition.

10.4.2. Materials

As shown in Fig. 16, in the concrete version of Ants and food, the ants were depicted by relatively simple line drawings of black ants, and the food consisted of an orange peach and red apple. In the idealized version of Ants and food, the ants were small black dots and the food sources were green patches. Participants "painted" food in the same manner in the concrete and idealized simulations, although in the concrete version, fractional portions of food were not allowed (see Fig. 16). The instructions for the two versions of Ants and food were changed to reflect their different appearances, and extra measures were taken to assure that participants would interpret the small black dots as ants. In particular, in a diagram from the simulation, green and black dots were labeled as food and ants, respectively.

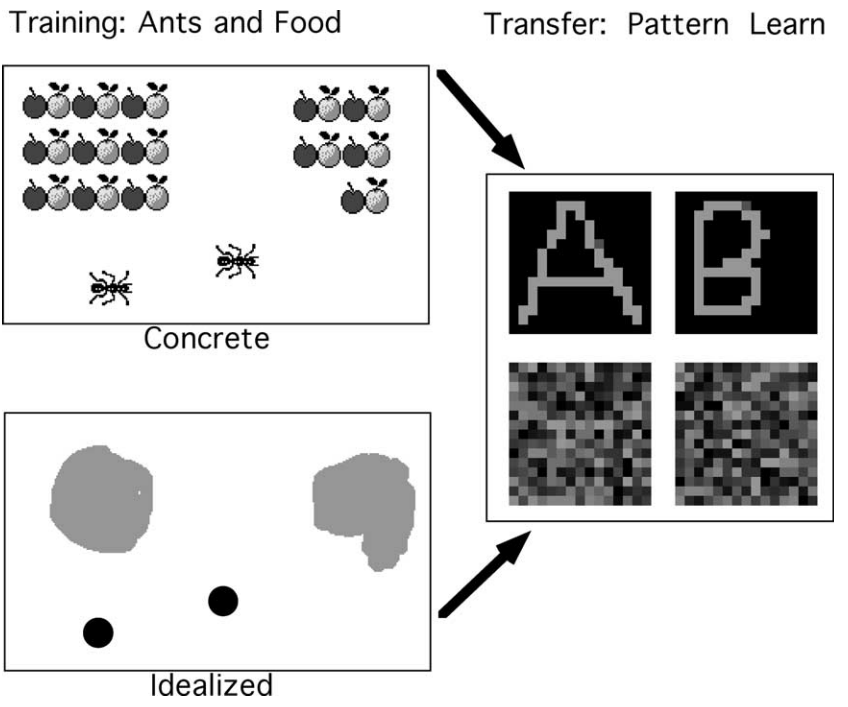

Fig. 16. The design for Experiment 4. The elements of the initial "Ants and food" simulation were either depicted with relatively concrete graphics that transparently revealed the entities that they represented, or were depicted with simple, idealized graphics.

10.4.3. Procedure

As illustrated in Fig. 16, the concrete and idealized conditions were always trained on Ants and food, and then were transferred to Pattern learn. At the end of each of the 20-min exploration periods, participants were given quizzes (see Appendix A). After the second quiz, they were given the analogical correspondences quiz. When time permitted, at the end of the experiment, participants were interviewed about the strategies they used to interact with the simulations, and about their understanding of how the agents within a simulation behaved. Only 12 out of 61 participants were interviewed.

10.4.4. Results

Concrete and idealized simulations differed in their quiz performances on the Ants and food simulation, with average scores of 39.8 and 33.8%, respectively, unpaired $t(59) = 3.9$, $p < .01$. Despite the significantly better performance on the concrete simulation during the initial simulation, there was a significant trend for the idealized simulation to produce better transfer to Pattern learn (41.3%) than did the concrete simulation (36.4%), unpaired $t(59) = 2.3$, $p < .05$. On the quiz testing knowledge of analogous elements across the simulations, there was not a significant difference between idealized and concrete conditions, with respective accuracies of 46.3 and 44.8% unpaired $t(59) = 0.8$, $p > .1$.

The superior transfer performance for idealized relative to concrete graphics must be interpreted with caution because simulation concreteness was involved in a cross-over interaction with initial simulation performance level. Based on quiz performance on Ants and food, participants from each concreteness condition were divided into two groups based on whether a participant got fewer than three quiz answers correct (poor performers) or more than two answers correct (good performers). The number of participants in good-concrete, good-idealized, poor-concrete, and poor-idealized conditions were 13, 15, 15, and 17. Collapsing over concrete and idealized conditions, the poor and good performers had mean initial quiz performances of 25.1 and 48.7%, respectively. There was an interaction between simulation concreteness and first simulation performance with the second, transfer simulation as the dependent measure, $F(1,57) = 10.8$, $p < .01$. As shown in Fig. 17, for poor initial simulation performers, the idealized simulation produced better transfer than the concrete simulation. For good performers, exactly the opposite pattern was obtained. Post hoc analyses revealed that these differences were significant at $p < .001$ and $p < .05$, respectively. There was not a significant interaction between similarity and first quiz performance on performance in the analogy quiz, $F(1,57) = 2.0$, $p > .1$.

We defined a competitive specialization strategy as any situation where a participant set the amount of adjustment for the closest category/ant to be much higher than that for the other categories/ants, but set both values to be greater than 0. We selected a difference of 40 as the threshold for signaling a competitive specialization strategy. As with the simulated annealing strategy measure, all key presses not directed toward adjusting the amount of adjustment of the closest category/ant and other categories/ants were ignored. The frequency of using a competitive specializa-

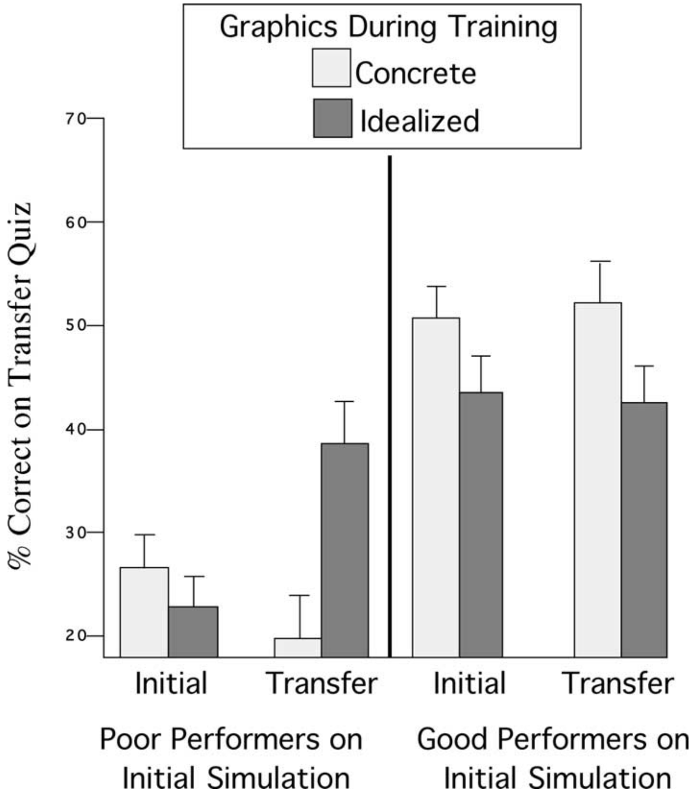

Fig. 17. Quiz results from Experiment 4. Participants performed better on the initial simulation when it contained relatively concrete, rather than idealized, graphical elements. However, for participants that performed relatively poorly on this initial simulation, transfer was better when the initial simulation contained idealized, rather than concrete, graphical elements.

tion response strategy was correlated with quiz performance for both the initial simulation (Pearson's $r = .24$, $p < .01$) and the transfer simulation ($r = .26$, $p < .01$). As shown in Fig. 18, good performers on the initial quiz used a greater number of competitive specialization response strategies than poor performers, on both the initial simulation, unpaired $t(59) = 4.1$, $p < .01$, and the transfer simulation, unpaired $t(59) = 3.5$, $p < .01$. There was not a significantly greater number of competitive specialization response strategies for the transfer than initial simulation, paired $t(60) = 1.1$, $p > .1$. There also was not a significant interaction between concreteness and first simulation quiz performance with number of competitive specialization responses on the transfer simulation as the dependent variable, $F(1, 57) = 1.5$, $p > .1$. Consistent with the results for initial quiz performance, use of the simulated competitive specialization strategy was greater for participants in the concrete rather than idealized condition, unpaired $t(59) = 2.3$, $p < .05$.

Unlike Experiments 1–3, the surface-level appearances manipulation had an impact on initial simulation performance. Participants performed better on the Ants

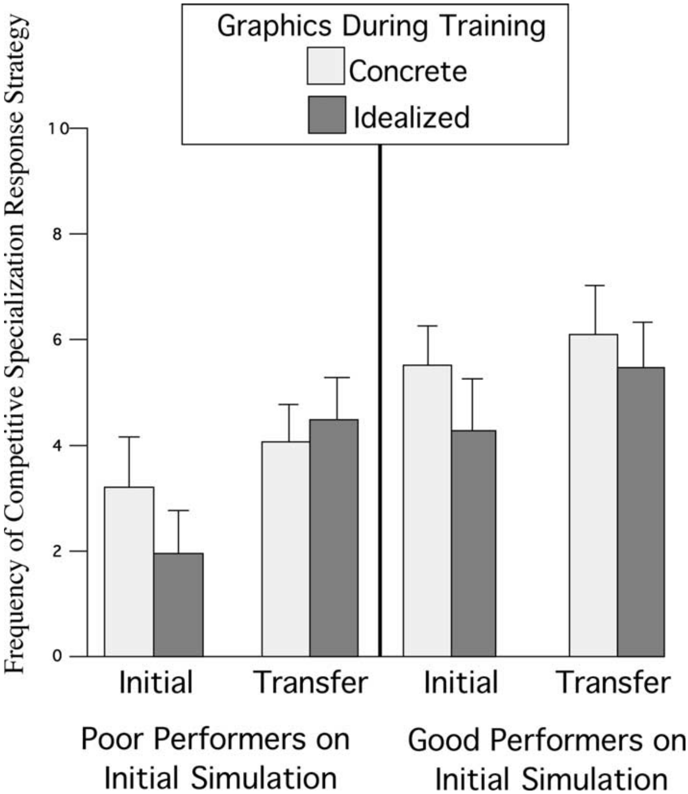

Fig. 18. Competitive specialization response strategies from Experiment 4.

and food quiz when the ants were easily identifiable as ants rather than appearing as abstract dots. Despite the better performance of the concrete graphics group on the concrete simulation itself, this group showed less successful transfer to another simulation that involved the same competitive specialization principle. Thus, there is a dissociation between performance on a task and generalization of the performance to an analogous task. This dissociation can be interpreted in terms of how tightly a learner's knowledge is tied to the particular training domain. If a learner's knowledge is tightly tied to a domain, then the learner may be able to perform well in this domain, but may show little ability to transfer their knowledge to related domains. By contrast, if the learner's knowledge is more abstract, it will, by definition, transfer well to analogous domains, but this increased capacity for transfer may be at the expense of a solid and concrete understanding of the original domain.

The results from the informal interviews must be interpreted with caution because only 12 participants were interviewed, and the responses are open-ended. Bearing in mind these limitations, the interviews did suggest that participants in the idealized graphics condition thought less about the particular constraints that an ant might face when approaching food, and thought more about the domain-general principles of competition between agents trying to cover resources. For example, learners in the concrete condition were more likely than learners in the idealized condition to describe the ants as "getting tired," "being happy with their share of food," "seeing another ant already occupying a food patch," and "scaring other ants away." In a sense, all of these descriptions are at odds with the rules explicitly given as describing the ants' movements. However, a learner who developed these anthropocentric (more accurately, arthropocentric) descriptions of the ants' behavior may have been able to develop easily comprehended accounts of the Ants and food simulation that were reasonably well correlated with the correct account. For example, although ants do not literally scare each other away, this act would produce similar behavior to the competitive specialization principle that all ants that are not closest to a selected food patch move very slowly toward the patch. The arthropocentric interpretations of the ants' behavior may make (properties correlated with) competitive specialization principles more understandable, and relatively realistic renderings of ants seems to predispose learners toward these interpretations. However, once these domain-specific interpretations are made, learners' understanding does not transfer very effectively to new competitive specialization situations.

The above discussion of the superior transfer from idealized, relative to concrete, simulations requires a caveat. Namely, this main effect is qualified by a large interaction involving learners' performance on the initial simulation. Relatively poor performers transfer better with the idealized simulation. Relatively good performers transfer better with the concrete simulation. The main effect of concreteness is found despite this interaction because the former difference is much larger than the latter difference. The interaction is not consistent with the theory that poor performers are in relative need of superficial details to extract and transfer abstract principles. Instead, the interaction suggests that poor performers benefit from idealized representations. Consistent with the accounts presented for Experiments 1–3, poor performers may be relatively susceptible to superficial properties of a simulation that distract a learner from the abstract principles.

11. General discussion

Four experiments explored transfer between simulations that were governed by the same abstract principle. Experiments 1–3 found that the transfer between abstractly related domains was modulated by the similarity between the domains for participants with relatively poor comprehension of the original simulation. In these experiments, greater positive transfer was found when superficial similarity was relatively low for poor comprehenders. This was true when overall superficial similarity was manipulated (Experiments 2–3), and when it was equated between high and low similarity conditions, with only similarity between analogous simulation entities being manipulated (Experiment 1). Although similarity manipulations had an effect on transfer, they did not systematically affect poor comprehenders' explicit understanding of the analogy relations between simulations as measured by their ability to correctly match corresponding entities across domains.

While Experiments 1–3 suggested a disadvantage of high cross-domain superficial similarity with respect to transfer for poor comprehenders, Experiment 4 found another situation in which superficial elements interfere with transfer. Better transfer was found for poor comprehenders from an initial to transfer simulation when the initial simulation was presented with simplified and idealized graphical elements rather than relatively concrete and clearly pictorial elements. The transfer advantage for idealized graphics was surprising given that participants in the concrete elements condition showed better understanding of the original domain. Experiment 4 suggested that understanding of a domain can be dissociated from transfer of that domain's knowledge. One account for this dissociation is that participants in the concrete elements condition were able to successfully answer the quiz questions that were designed to probe abstract knowledge by using domain-specific reasoning that was sufficiently related to the abstract principles probed by the quiz. Although this domain-specific understanding could successfully ground a partially abstract understanding of the domain, it would curtail transfer to new domains. Conversely, idealized graphics might introduce difficulties in comprehending the initial domain if they weakened the intuitive connection between graphical elements and their interpretations, but whatever partial understanding is achieved would be expected to transfer well if it is less tied to the particulars of the domain.

Each of the experiments found an interaction between the manipulation of superficial appearances and initial performance level, with transfer performance as the dependent measure. In Experiments 1–3, poor initial performers showed the largest influence of superficial similarity. Good performers on the initial simulation performed about equally well on the transfer simulation with superficially similar and dissimilar domains. Poor performers transferred much better with the dissimilar, relative to similar, domains. Poor performers seem particularly adversely affected by high superficial similarity between domains. Likewise, in Experiment 4, poor performers were particularly adversely affected by concrete, relative to idealized, simulations. In fact, in Experiment 4, the opposite effect was found for good performers, with concrete simulations transferring better than idealized simulations. Rather than viewing poor performers as in particular need of superficial properties and similarities to help them extract a deep understanding, Experiments 1–3 indicate that poor performers are particularly at risk for being mislead by superficial properties and similarities.

The experiments bear on explicit and implicit transfer of knowledge. One of the measures of transfer used by all of the experiments was performance on a multiple-choice quiz testing understanding of an abstract principle in a specific domain. This is a relatively explicit measure of a learner's knowledge, requiring them to identify the correct written description of the behavior of a simulation. Implicit measures were also developed in all of the experiments by observing the participants' interactive usage of the transfer simulation. Key presses diagnostic of using simulated annealing and competitive specialization strategies were identified. In Experiments 2 and 3, the implicit measures showed the same interaction between learner's initial performance level and similarity as was shown by the quiz-based measure of transfer. Experiment 3 also explored an implicit measure of performance based on the ability of learners to solve problems in a domain. Solution time to solve a transfer problem was affected in the same way as the other implicit measure by the similarity between the problems and the initial performance level of participants. These are important results because they show that the observed transfer effects are not solely attributable to repetitions of the same kind of quiz problems across two simulations. Moreover, the implicit measure of transfer shows that the initial simulation has an influence on participants even while they are interacting with the transfer simulation.

11.1. Facilitation and interference between superficial and abstract properties

The experiments bear directly on the nature of the relation between apprehending abstract and superficial properties of a situation. As described in Section 4, there are reasons for thinking that superficial features may either facilitate or interfere with abstract comprehension. On the one hand, recognizing superficial similarities between two situations may bootstrap the formation of more abstract similarities between them (Gentner & Wolff, 2000; Kotovsky & Gentner, 1996; Mix, 1999; Ross, 1984, 1989, 1990). One straightforward way for this to occur is if superficial similarities between two situations remind one of the earlier situation when the later situation is introduced (Gick & Holyoak, 1980; Ross et al., 1990). Once the two situations are explicitly compared to each other because of the reminding, more abstract commonalities between them may be noticed (Ross & Kennedy, 1990).

On the other hand, other evidence suggests that superficial and abstract understandings may compete against each other (Goldstone et al., 1991; DeLoache, 1991, 1995; Uttal et al., 1999). If the superficial properties of a situation, or the superficial similarities between situations, are too salient, then people sometimes show diminished sensitivity to the abstract properties. Goldstone et al. (1991) described this competition in terms of separate pools for abstract, relational similarities and for superficial similarities, arguing that as one pool gets larger, it attracts attention toward itself and away from the other pool. DeLoache theorizes that there exists a competition between viewing an object as a concrete object in its own right, and as symbol for understanding another object. Increasing the salience of the concrete properties of an object interferes with its use as a symbol because of this competition.

The experimental results support the competitive, rather than facilitative, theories of the relation between appreciating superficial and abstract properties. However, it would be premature to broadly generalize this conclusion, particularly given the large and reasonably consistent literature suggesting that superficial similarities between analogous domains increase the likelihood of transferring abstract solutions between them (Reeves & Weisberg, 1994; Spencer & Weisberg, 1986). Instead, a more reasonable approach is to try to understand why the current experimental design shows competition when many other paradigms have shown facilitation. One promising account for the difference between the current results and many previous results is that the advantage of high superficial similarity in terms of promoting remindings may have not been very important in the current paradigm. When people are asked to solve multiple mathematical word problems (Ross, 1987, 1988) or insight problems (Gick & Holyoak, 1980, 1983; Novick & Holyoak, 1991; Reed & Bolstad, 1991), there is little a priori reason to think the problems will be related to each other. If participants are not given a hint that previous solutions are useful in solving a problem, then it may not naturally occur for participants to consider the previous problems. In this context, increasing the superficial similarity between the problems could spur participants to remember the previous problem and its solution.

In contrast, several aspects of the current experiments would have encouraged participants to be spontaneously reminded of the earlier simulation while exploring the later simulation. There was no delay between the two simulations, the simulations' screen layouts were highly similar, the abstract descriptions of the rules underlying analogous simulations were very similar, and analogous simulation quantities were plotted in the same manner in the continuously updated graph. For these reasons, participants were likely to see some link between the two simulations even when they did not appreciate the abstract principle that linked them together. This argument is supported by the questionnaire concerning remindings given at the end of Experiment 3. Participants indicated a high spontaneous rate of remindings. In this kind of situation, one of the principal benefits of high superficial similarity would not have been needed, and the cost associated with high superficial similarity could become evident.

One useful framework for reconciling the costs and advantages of high cross-domain superficial similarity is Medin and Ross' notion of conservative generalization (Medin & Ross, 1989; Ross & Kennedy, 1990). According to this notion, the generalization that is mentally created to summarize several domains will emphasize all of the salient features held in common by the domains, including both superficial features and more abstract properties. If the domains are overall very similar, then the generalization will have a relatively low ratio of abstract to superficial features, and will not generalize well to new domains with different superficial appearances. If one further assumes that abstract features are drowned out by a large number of other features within a generalization and hence tend to be ignored, then the decreased abstraction-based transfer with increased superficial similarity observed in Experiments 1 and 2 is predicted. In short, superficial similarity increases the likelihood that a cross-domain generalization will be attempted, but decreases the abstractness of this representation (see also Gick & Holyoak, 1983).

11.2. Pedagogical implications

One of the assumptions underlying the computer simulations developed here is that the abstract principles underlying complex adaptive systems are effectively conveyed by concrete, graphical, interactive, and dynamic simulations (Resnick, 1994; Resnick & Wilensky, 1998). Perceptual processes can be used to bootstrap abstract concepts that would otherwise be difficult to convey (Goldstone & Barsalou, 1998). Even granting this assumption, there are many unanswered empirical questions related to the optimal level of concreteness and the relation between original and transfer domains that optimizes transfer. Within educational psychology there is an active debate as to the level of abstract knowledge that teachers should aspire to instill in their students (Vera & Simon, 1993). According to a theoretical position called "situated learning," much of what is learned is specific to the situation in which it was learned (Lave, 1988).[2] This view predicts that knowledge is grounded in the concrete, real-world context in which it occurs, and as such often does not transfer to different contexts, particularly those involving artificial scholastic or laboratory settings. This position is often also associated with the claim that training by presenting abstractions is of little use, and that training should involve authentic, real-world problems (Brown, Collins, & Duguid, 1989; Collins, Brown, & Newman, 1989). However, other researchers have taken issue with these claims, arguing that transfer based on abstractions does occur, that real-world problems do not always efficiently highlight important skills that can be exported to new situations, and that teaching abstract systems is often an effective way of promoting generally applicable knowledge (Anderson et al., 1996).

A small set of experiments cannot resolve these two conflicting theoretical perspectives, but the present experiments do bear on the debate on the proper role of abstractions in education. The pilot experiment to Experiment 1 provided evidence for better transfer between simulations that are abstractly related rather than unrelated simulations. Such a result indicates that what is learned during interactive exploration of a concrete simulation is not completely tied to the simulation's concrete domain. For example, learning about ants foraging for food transfers to categories adapting to pictures because they are both governed by the principle of competitive specialization.

An additional result from the experiments was that increasing the concreteness of a simulation does not always make it more effective educationally. Although concrete, relative to idealized, simulations resulted in better performance with respect to a quiz testing abstract knowledge of the simulation itself, overall transfer was better with the more idealized simulation. For learners who showed originally poor knowledge of the initial simulation's abstract principle, the idealized simulation resulted in far more effective transfer. For good performers, the opposite trend was found. If this kind of result continues to be found, several pedagogical implications would follow. First, increasing the realism of a tutorial system does not guarantee that transferable knowledge is increased. Second, performance on the tutorial system itself may be a misleading gauge as to how well the knowledge obtained will transfer. Third, different levels of realism may be beneficial for different learners depending on their achievement level. In contrast to some claims made by situated learning theorists, the experiments speak to the beneficial impact of making simulations less concrete, and decreasing the superficial similarity between initial and transfer simulations.

This notion, that too much situational detail can impair abstract understanding and therefore transfer, is prevalent among math educators who design highly generic word problems with few salient details to distract students (Kintsch & Greeno, 1985;

[2] Lave's claim for very limited transfer to novel contexts is not shared by all situated learning theorists (e.g., Greeno, Smith, & Moore, 1993), and some argue that this is not in fact Lave's position (Greeno, 1997). However, other authors (Anderson et al., 1996; Cox, 1997) have interpreted Lave (1998) as making this claim

Nesher & Katriel, 1977). Most word problems contain only minimal information about situations, in part to help students align situations with underlying mathematical structures (Bassok, 2001). Others criticize this practice, instead recommending math instruction that employs more realistic situations (Barron et al., 1998). Although more experimentation is certainly needed, our results indicate that decreasing the strength of the link between the real world and its model is beneficial in some occasions.

12. Conclusions

A variety of experiments suggest that perceptually concrete properties and superficial similarities are helpful in building up and transferring abstract knowledge. However, the current experiments present some exceptions to this generalization. Superficial similarities between scenarios do not universally promote abstraction-based transfer. They also inhibit this transfer for relatively poor comprehenders of an initial situation. Concrete graphical elements do not universally promote abstraction-based transfer. They can also inhibit this transfer, at least for learners who do not originally show strong evidence of comprehending the abstraction. Both results can be explained in terms of a competition between interpreting a scenario in an abstract versus concrete manner. Increasing concreteness and surface-level similarity can distract a learner from taking a more abstract perspective, particularly if the learner is already prone to miss abstractions. Poor comprehenders of an abstract principle may need less, rather than more, realism to understand and transfer the principle.

Given the increasing importance of computer simulations as a vehicle for teaching difficult, abstract concepts, there is a profound need for further research on how these simulations should be presented. The current studies underscore how little we know about the processes by which concrete simulations facilitate abstract understanding. Contrary to a common assumption of virtual reality research, greater concreteness and model-world transparency do not universally promote abstract understanding. Contrary to expectations derived from problem solving, increased surface-level similarity does not universally promote abstraction-based transfer. Determining what are the costs and benefits of emphasizing concrete properties for conveying abstract principles could go a long way toward improving our educational methods.

Acknowledgments

Many useful comments and suggestions were provided by Miriam Bassok, Judy DeLoache, Kevin Dunbar, Dedre Gentner, Douglas Medin, Kelly Mix, Laura Novick, Steven Reed, and Jeffrey Reynolds. We would like to thank Melissa Adkins, Adam Fritz, and Brent Myers for assistance in running the experiments. This research was funded by NIH Grant MH56871, and NSF Grant 0125287.

Appendix A. Sample multiple-choice quiz questions for simulations

Correct answers are indicated with asterisks. Rather than showing all seven questions for each simulation, two questions are shown with their analogs and two questions are shown without their analogs.

A.1. Dropping balls

If the aim is to get the balls to the lowest point, why does adding random movement to the balls help?
1. Random balls are more likely to go beyond how they were programmed.
2. Randomness allows the balls to avoid the shallow valleys more easily because the shallow valleys are in fixed locations.
3. Randomness allows the balls to move unpredictably, and this allows them to surprise us with good solutions.
*4. Randomness allows the balls to escape shallow valleys that would otherwise trap them.

If the aim is to get the balls to the lowest point, what is the problem with having balls move mostly at random?
1. Randomness can never produce a good solution because it leaves everything up to chance.
*2. Randomly moving balls will never completely settle down into a low valley.
3. The balls will settle on a good solution at the end, but will take a long time to get there.
4. The balls will not be predictable in their behavior, so we cannot say what the solution will look like.

What is the best method for getting balls to reach a low point?
*1. Start out with a lot of randomness, and then gradually decrease the randomness.
2. Start out with very little randomness, and then gradually increase the randomness.
3. Keep a consistently low level of randomness.
4. Keep a consistently high level of randomness.

Sometimes balls get stuck in shallow valleys. What can be done to get them out of these valleys?
*1. Increase the randomness of the balls' movements, and make them move a lot.
2. Decrease the randomness and make the balls move by only a small amount.
3. Make the balls move a large amount, but keep their randomness low.
4. Decrease the number of balls that are dropping.

A.2. Path finder

When trying to find a pathway between two endpoints avoiding obstacles, why does adding random movement to the balls help?
1. Random balls are more likely to go beyond how they were programmed.
2. Randomness allows the balls to avoid the obstacles more easily because the obstacles are in fixed locations.
3. Randomness allows each ball to move unpredictably, which allows them to surprise us with good solutions.
*4. Randomness allows the balls to break out of patterns that are not moving towards a good pathway.

If the aim is to get the balls to form a good pathway, what is the problem with having balls moving mostly at random?
1. Randomness can never produce a good solution because it leaves everything up to chance.
*2. Randomly moving balls will never completely settle down into a single pathway.
3. The balls will settle on a good solution at the end, but will take a long time to get there.
4. The balls will not be predictable in their behavior, so we cannot say what the solution will look like.

If there is no randomness to the balls' movements, then they will always move towards their neighbors (unless there is an obstacle). Does this always lead to the lowest possible distance between neighboring balls?

1. Yes, because each ball will move exactly in the optimal way for positioning itself between its two neighbors.
2. No, because moving toward one neighbor necessarily moves the ball away from another neighbor.
3. No, because the neighboring balls that a ball moves toward are themselves moving, and may not be where they were before.
*4. No, because they might have been able to be even closer if they had tried a very different configuration, but they cannot find it because of the low randomness.

The best way to have balls find good paths is by gradually reducing the randomness in their movements? Why?

1. As the noisy influence of the randomness is gradually eliminated, the balls more easily avoid the obstacles in their way.
2. The noisy influence of the randomness is gradually eliminated, allowing the balls to follow their neighbors.
*3. Early on, randomness helps the balls break out of bad solutions. By eliminating randomness later on, the balls settle down into a single pathway.
4. Randomness allows the balls to explore different pathways, and then when it is removed, the balls return to the best path that they found earlier.

A.3. Ants and food

To make the ants as a population cover the food well, which strategy is the best:

1. Have the ants move as quickly as possible.
*2. Make the ant that is closest to a piece of food move more quickly than all the other ants.
3. Make the ant that is closest to a piece of food move more slowly than all the other ants.
4. Early on, make the closest ant move more quickly than the others, but later on, make the closest ant move more slowly.

Why don't the ants cover the food well if only the ant closest to a piece of food moves?

*1. If only the closest ant moves, then this ant may be responsible for many pieces of food, and the other ants may not cover any food.
2. If only the closest ant moves, then it may eventually get tired and stop moving at all.
3. The ant closest to a piece of food shouldn't move much because it is already close to the food. It is the other ants that need to move.
4. If only one ant moves, then as a population, the ants are not moving very much when a piece of food is selected.

Why don't the ants cover the food well if the closest ant and all of the other ants all move with the same speed?

1. The closest ant doesn't need to move as quickly as the other ants because it is already close to the food.
*2. If other ants move as fast as the closest ant, then when a new piece of food is selected, they will move away from food that they were previously covering well.
3. If all of the ants move with the same speed, then they will all get to the food at the same time, and they won't be able to share it efficiently if there are a lot of ants.
4. If the closest ant moves as fast as the other ants, then it will get to the food first, and will prevent the other ants from benefiting from it.

To have the ants cover the food well, it is necessary to have the ants become specialized for particular food patches. Which action most directly allows for this specialization?

1. Make sure that there are not very many ants on the field. That way, each ant can be far away from other ants.
2. Make sure that there are many ants on the field. That way, each ant can become specialized for a tiny patch.
*3. Make the ants that are not closest to a piece of food move slowly to the food.
4. Make the ant that is closest to a piece of food move slowly to the food.

A.4. Pattern learn

To make categories best represent the natural groups in a set of pictures, you should:
1. Have the categories adapt as quickly as possible.
*2. Make the category that is closest to a selected picture adapt more quickly than all the other categories.
3. Make the category that is closest to a selected picture adapt more slowly than all the other categories.
4. Early on, make the closest category adapt itself more quickly than the others, but later on, make the closest category adapt more slowly.

Why aren't good categories formed if only the picture closest to a selected picture adapts?
*1. If only the closest category adapts, then this category will become responsible for many pictures, and the other categories may not be adapted for any picture.
2. If only the closest category adapts, then its learning may eventually become exhausted, and it may stop learning at all.
3. The picture closest to a picture shouldn't adapt much because it is already close to the picture. It is the other pictures that need to adapt.
4. If only one category adapts, then in general the categories are not changing very much when a picture is selected.

If there are four pictures and two categories, the categories will emphasize the parts that the pictures in a category share. How does this occur?
1. Categories adapt most toward parts shared by category members because shared parts provide links between pictures belonging to different categories.
2. Categories adapt most toward parts shared by category members because the rate of adaptation will be faster for categories that are not closest to the selected picture.
*3. Categories adapt most toward parts shared by category members because these parts are always present in the members, and so there will be more opportunities for learning.
4. Categories adapt most toward parts shared by category members because these parts are at the very essence of the category, defining what it means to be part of the category.

If there are two pictures and only one category, what usually happens?
1. The category will alternate between the pictures, but only if it adapts very slowly.
*2. The category will be a blend of the two pictures, highlighting parts shared by the pictures.
3. The category will become specialized for one of the pictures only.
4. The category will not become adapted to either picture, unless it is highly similar to them in the first place.

Appendix B. Sample multiple-choice quiz questions for describing analogies between simulations

Correct answers are indicated with asterisks.

In "Path finder," each ball's goal was to have close neighboring balls. What was the analogous goal in "Dropping balls"?:
1. Balls moved so that they would form clusters.
2. Balls moved so that they would form paths between valleys.
*3. Balls moved to the lowest possible spot.
4. Balls moved so as to be as close as possible to other balls.

In "Dropping balls," balls sometimes fell into shallow valleys rather than the deepest valley. What was the analogous phenomenon in "Path finder"?
1. Balls sometimes fell to the bottom of the screen without forming a good pathway.
*2. Balls sometimes remained far away from their neighbors, separated by an obstacle.
3. Balls sometimes moved randomly around without forming a good pathway.
4. Balls sometimes moved away from other balls when moving randomly.

In “Path finder,” balls sometimes found arrangements where each was very close to its neighbor. What is the analogous phenomenon in “Balls Dropping”?

*1. Balls sometimes escaped from shallow valleys, falling into lower ones.
2. Balls sometimes formed clusters.
3. Balls sometimes arranged themselves in patterns while dropping.
4. Different balls sometimes fell on to the same spot.

In “Dropping balls,” randomness allowed the balls to move upwards, out of shallow valleys. What was the analogous thing that randomness allowed in “Path finder”?

1. It allowed the balls to wiggle around the screen.
2. It allowed the balls to move away from the obstacles.
*3. It allowed the balls to not always move directly toward their neighbors.
4. It allowed the balls to not always move downwards.

References

Anderson, J. R., Reder, L. M., & Simon, H. A. (1996). Situated learning and education. *Educational Researcher, 25*, 5–11.

Ball, P. (1999). *The self-made tapestry*. Oxford, England: Oxford University Press.

Bak, P. (1996). *How nature works: The science of self-organized criticality*. New York: Springer.

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

Barron, B. J. S., Schwartz, D. L., Vye, N. J., Moore, A., Petrosino, A., Zech, L., & Bransford, J. D. (1998). Doing with understanding: lessons from research on problem- and project-based learning. *Journal of Learning Sciences, 7*, 271–311.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577–660.

Bassok, M. (1996). Using content to interpret structure: Effects on analogical transfer. *Current Directions in Psychological Science, 5*, 54–58.

Bassok, M. (2001). Semantic alignments in mathematical word problems. In D. Gentner, K. J. Holyoak, & B. Kokinov (Eds.), *The analogical mind: Perspectives from cognitive science* (pp. 401–433). Cambridge, MA: MIT Press.

Bassok, M., Chase, V. M., & Martin, S. A. (1998). Adding apples and oranges: Alignment of semantic and formal knowledge. *Cognitive Psychology, 35*, 99–134.

Bassok, M., & Holyoak, K. J. (1989). Interdomain transfer between isomorphic topics in algebra and physics. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 153–166.

Bassok, M., & Olseth, K. L. (1995). Object-based representations: Transfer between cases of continuous and discrete models of change. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1522–1538.

Bassok, M., Wu, L., & Olseth, K. L. (1995). Judging a book by its cover: Interpretative effects of content on problem-solving transfer. *Memory & Cognition, 23*, 354–367.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*, 34–41.

Campione, J. C. (1985). Breakdowns in flexible use of information: Intelligence-related differences in transfer following equivalent learning performance. *Intelligence, 9*, 297–315.

Casti, J. L. (1994). *Complexification*. New York: HarperCollins.

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

Cheng, P. C. H. (2002). Electrifying diagrams for learning: Principles for complex representational systems. *Cognitive Science, 26*, 685–736.

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

Collins, A., Brown, J. S., & Newman, S. (1989). Cognitive apprenticeship: Teaching students the craft of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honors of Robert Glaser* (pp. 453–494). Hillsdale, NJ: Erlbaum.

Cox, B. D. (1997). The rediscovery of the active learner in adaptive contexts: A developmental–historical analysis of the transfer of training. *Educational Psychologist, 32*, 41–55.

Darwin, C. (1859). *On the origin of species by means of natural selection*. London: Murray.

DeLoache, J. S. (1991). Symbolic functioning in very young children: Understanding of pictures and models. *Child Development, 62*, 736–752.

DeLoache, J. S. (1995). Early understanding and use of symbols: The model model. *Current Directions in Psychological Science, 4*, 109–113.

DeLoache, J. S., & Burns, N. M. (1994). Symbolic functioning in preschool children. *Journal of Applied Developmental Psychology, 15*, 513–527.

DeLoache, J. S., & Marzolf, D. P. (1992). When a picture is not worth a thousand words: Young children's understanding of pictures and models. *Cognitive Development, 7*, 317–329.

DiFonzo, N., Hantula, D. A., & Bordia, P. (1998). Microworlds for experimental research: Having your (control and collections) cake, and realism too. *Behavioral Research Methods Instruments, and Computers, 30*, 278–286.

Einstein, A. (1989). *The collected papers of Albert Einstein. Vol. 2: The Swiss years: writings, 1900–1909*. Princeton, NJ: Princeton University Press.

Flake, G. W. (1998). *The computational beauty of nature*. Cambridge, MA: MIT Press.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7*, 155–170.

Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou & A. Ortony (Eds.), *Similarity, analogy, and thought* (pp. 199–241). New York: Cambridge University Press.

Gentner, D., & Stevens, A. L. (Eds.). (1983). *Mental models*. Hillsdale, NJ: Erlbaum.

Gentner, D., & Toupin, C. (1986). Systematicity and surface similarity in the development of analogy. *Cognitive Science, 10*(3), 277–300.

Gentner, D., & Wolff, P. (2000). Metaphor and knowledge change. In E. Dietrich & A. B. Markman (Eds.), *Cognitive dynamics: Conceptual and representational change in humans and machines* (pp. 294–342). Mahwah, NJ: Lawrence Erlbaum.

Gianutsos, R. (1994). Driving advisement with the elemental driving simulator (EDS): When less suffices. *Behavioral Research Methods, Instruments, & Computers, 26*, 183–186.

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*, 1–39.

Gilden, D. L., Thornton, T., & Mallon, M. W. (1995). $1/f$ noise in human cognition. *Science, 267*, 1837–1839.

Goldstone, R. L. (1994a). The role of similarity in categorization: Providing a groundwork. *Cognition, 52*, 125–157.

Goldstone, R. L. (1994b). Similarity, interactive activation, and mapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 3–28.

Goldstone R. L. (in press). Learning to perceive while perceiving to learn. In: R. Kimchi, M. Behrmann, & C. Olson (Eds.), *Perceptual organization in vision: Behavioral and neural perspectives*.

Goldstone, R. L., & Barsalou, L. (1998). Reuniting perception and conception. *Cognition, 65*, 231–262.

Goldstone, R. L., & Medin, D. L. (1994). The time course of comparison. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 29–50.

Goldstone, R. L., Medin, D. L., & Gentner, D. (1991). Relational similarity and the nonindependence of features in similarity judgments. *Cognitive Psychology, 23*, 222–264.

Grady, S. M. (1998). *Virtual reality: Computers mimic the physical world*. New York: Facts on File.

Greeno, J. (1997). Response: On claims that answer the wrong questions. *Educational Researcher, 26*, 5–17.

Greeno, J. G., Smith, D. R., & Moore, J. L. (1993). Transfer of situated learning. In D. Detterman & R. J. Sternberg (Eds.), *Transfer on trial: Intelligence, cognition, and instruction* (pp. 99–167). Norwood, NJ: Ablex.

Grossberg, S. (1976). Adaptive pattern classification and universal recoding I: Parallel development and coding of neural feature detectors. *Biological Cybernetics, 23*, 121–134.
Heim, M. (2000). *Virtual realism*. Oxford: Oxford University Press.
Holyoak, K. J. (1984). Analogical thinking and human intelligence. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 2, pp. 199–230). Hillsdale, NJ: Erlbaum.
Holyoak, K. J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory & Cognition, 15*, 332–340.
Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review, 104*, 427–466.
Jacobson, M. J. (2000). Problem solving about complex systems: Differences between experts and novices. In B. J. Fishman & S. F. O'Connor-Divelbiss (Eds.), *Proceedings of the international conference of the learning sciences 2000* (pp. 14–21). Mahwah, NJ: Lawrence Erlbaum.
Jackson, S., Stratford, J., Krajcik, S., & Soloway, E. (1996). Making system dynamics modeling accessible to pre-college science students. *Interactive Learning Environments, 4*, 233–257.
Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, MA: Harvard University Press.
Keane, M. (1987). On retrieving analogues when solving problems. *Quarterly Journal of Experimental Psychology, 39A*, 29–41.
Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological Review, 92*, 109–129.
Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science, 220*, 671–680.
Klahr, D., & Carver, S. M. (1988). Cognitive objectives in a Logo debugging curriculum: Instruction, learning, and transfer. *Cognitive Psychology, 20*, 362–404.
Kohonen, T. (1995). *Self-organizing maps*. Berlin: Springer.
Kotovsky, L., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development, 67*, 2797–2822.
Lave, J. (1988). *Cognition in practice: Mind, mathematics, and culture in everyday life*. New York: Cambridge University Press.
Lovett, M. C., & Anderson, J. R. (1996). History of success and current context in problem solving: Combined influences on operator selection. *Cognitive Psychology, 31*, 168–217.
Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology, 25*, 431–467.
Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179–195). Cambridge: Cambridge University Press.
Medin, D. L., & Ross, B. H. (1989). The specific character of abstract thought: Categorization, problem-solving, and induction. In R. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 5, pp. 120–154). Hillsdale, NJ: Erlbaum.
Miller, C. S., Lehman, J. F., & Koedinger, K. R. (1999). Goals and learning in microworlds. *Cognitive Science, 23*, 305–336.
Miller, K. F., & Stigler, J. W. (1991). Meanings of skill: Effects of abacus expertise on number representation. *Cognition and Instruction, 8*, 29–67.
Mix, K. S. (1999). Similarity and numerical equivalence: Appearances counts. *Cognitive Development, 14*, 269–297.
Needham, D. R., & Begg, I. M. (1991). Problem-oriented training promotes spontaneous analogical transfer: Memory-oriented training promotes memory for training. *Memory & Cognition, 19*, 543–557.
Nesher, P., & Katriel, T. A. (1977). A semantic analysis of addition and subtraction word problems in arithmetic. *Educational Studies in Mathematics, 8*, 251–269.
Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.
Novick, L. R., & Holyoak, K. J. (1991). Mathematical problem solving by analogy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 398–415.
National Research Council (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

O'Reilly, R. C. (2001). Generalization in interactive networks: The benefits of inhibitory competition and Hebbian learning. *Neural Computation, 13*, 1199–1242.
Pedone, R., Hummel, J. E., & Holyoak, K. J. (2001). The use of diagrams in analogical problem solving. *Memory & Cognition, 29*, 214–221.
Ratterman, M. J., & Gentner, D. (1998). The effect of language on similarity: The use of relational labels improves young children's performance in a mapping task. In K. Holyoak, D. Gentner, & B. Kokinov (Eds.), *Advances in analogy research: Integration of theory and data from the cognitive, computational, and neural sciences* (pp. 274–282). Sophia: New Bulgarian University.
Reed, S. K. (1999). *Word problems: Research and curriculum reform*. Mahwah, NJ: Lawrence Erlbaum.
Reed, S. K., & Bolstad, C. (1991). Use of examples and procedures in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 753–766.
Reed, S. K., Cooke, J., & Jazo, L. (2002). Building complex solutions from simple solutions in the Animation Tutor: Task completion. *Mathematical Thinking & Learning, 4*, 315–336.
Reeves, L. M., & Weisberg, R. W. (1994). The role of content and abstract information in analogical transfer. *Psychological Bulletin, 115*, 381–400.
Resnick, M. R. (1994). *Turtles, termites, and traffic jams*. Cambridge, MA: MIT Press.
Resnick, M. R., & Wilensky, U. (1998). Diving into complexity: Developing probabilistic decentralized thinking through role-playing activities. *Journal of Learning Sciences, 7*, 153–172.
Reyes, R. M., Thompson, W. C., & Bower, G. H. (1980). Judgmental biases resulting from differing availability of arguments. *Journal of Personality and Social Psychology, 39*, 2–12.
Ross, B. H. (1984). Remindings and their effects in learning a cognitive skill. *Cognitive Psychology, 16*, 371–416.
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.
Ross, B. H. (1989). Distinguishing types of superficial similarities: Different effects on the access and use of earlier problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 456–468.
Ross, B. H. (1990). The access and use of relevant knowledge: A special case and general issues. In R. Freedle (Ed.), *Artificial intelligence and the future of testing* (pp. 173–211). Hillsdale, NJ: Lawrence Erlbaum.
Ross, B. H., & Kennedy, P. T. (1990). Generalizing from the use of earlier examples in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 42–55.
Ross, B. H., Perkins, S. J., & Tenpenny, P. L. (1990). Reminding-based category learning. *Cognitive Psychology, 22*, 460–492.
Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 33–58). Hillsdale, NJ: Erlbaum.
Rumelhart, D. E., & Zipser, D. (1985). Feature discovery by competitive learning. *Cognitive Science, 9*, 75–112.
Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 501–518.
Schank, R. C., & Farrel, R. (1988). Creativity in education: A standard for computer-based teaching. *Machine Mediated Learning, 2*, 175–194.
Schiano, D. J., Cooper, L. A., Glaser, R., & Zhang, H. C. (1989). Highs are to lows as experts are to novices: Individual differences in the representation and solution of standardized figural analogies. *Human Performance, 2*, 225–248.
Schunn, C. D., & Dunbar, K. (1996). Primary, analogy, and awareness in complex reasoning. *Memory & Cognition, 24*, 271–284.
Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237*, 1317–1323.
Spencer, R. M., & Weisberg, R. W. (1986). Context-dependent effects on analogical transfer. *Memory & Cognition, 14*, 442–449.
Stadler, M. A. (1989). On learning complex procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 1061–1069.
Stephens, D. W., & Krebs, J. R. (1987). *Foraging theory*. New Haven: Princeton University Press.

Underwood, G. D. M. (1996). *Implicit cognition*. New York: Oxford University Press.

Uttal, D. H., Liu, L. L., & DeLoache, J. S. (1999). Taking a hard look at concreteness: Do concrete objects help young children learn symbolic relations?. In C. S. Tamis-LeMonda (Ed.), *Child psychology: A handbook of contemporary issues* (pp. 177–192). Philadelphia, PA: Psychology Press.

Vera, A. H., & Simon, H. A. (1993). Situated action: A symbolic interpretation. *Cognitive Science, 17*, 7–48.

von der Malsburg, C. (1973). Self-organization or orientation-sensitive cells in the striate cortex. *Kybernetik, 15*, 85–100.

Voss, J. F., Greene, T. R., Post, T. A., & Penner, B. C. (1984). Problem solving skills in the social sciences. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research theory* (Vol. 17). New York: Academic Press.

White, B. Y. (1993). ThinkerTools: Causal models, conceptual change, and science education. *Cognition and Instruction, 10*, 1–100.

White, B. Y., & Fredericksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction, 16*, 3–118.