The Prehistory of Discovery: Precursors of Representational Change in Solving Gear System Problems

James A. Dixon
University of Connecticut

Ashley S. Bangert
University of Michigan

Microgenetic research has identified 2 different types of processes that produce representational change: theory revision and redescription. Both processes have been implicated as important sources of developmental change, but their relative status across development has not been addressed. The current study investigated whether (a) the process of representational change undergoes developmental change itself or (b) different processes occupy different niches in the course of knowledge acquisition. College, 3rd-, and 6th-grade students solved gear system problems over 2 sessions. For all grades, discovery of the physical principles of the gear system was consistent with theory revision, but discovery of a more sophisticated strategy, based on the alternating sequence of gears, was consistent with redescription. The results suggest that these processes may occupy different niches in the course of acquiring knowledge and that the processes are developmentally invariant across a broad age range.

Recent research in cognitive development has focused on the processes that underlie cognitive change. Children spontaneously change their representations of physical, social, and formal domains as well as their strategies for addressing problems posed in those domains. For example, Chen and Klahr (1999) reported that children’s representations of specific domains, such as springs and sinking, improved as a result of conducting experiments. Similarly, Alibali (1999) showed that children’s understanding of the mathematical concept of equivalence changed substantially after a very brief period of instruction. The processes that underlie these cognitive changes are believed to be a driving force in cognitive development. The small, local changes in representations produced by these processes eventually culminate in fundamental changes in conceptual structure (Case & Okamoto, 1996; Demetriou & Raftopoulos, 1999).

The recent development of microgenetic methods to study children’s cognition has yielded important new data about the characteristics of cognitive change (Kuhn, 1995). As a result of these microgenetic studies, two major types of processes have been proposed to explain representational change. The more widely researched of these processes is theory revision, in which children change their representations of how a particular domain works in response to disconfirmatory evidence. For example, Schauble (1996) showed that children proposed initial hypotheses about the factors that affected the speed of a boat. After children performed empirical tests that disconfirmed their initial hypotheses, many children modified their hypotheses to reflect the new result. Kuhn, Garcia-Mila, Zohar, and Andersen (1995) emphasized that the process of theory revision involves the metacognitive coordination of hypotheses, empirical tests, and evidence:

If knowledge acquisition is a process of theory revision, as we have claimed, to accomplish the process in a skilled way the individual needs to be aware of and reflect on a theory (metacognitive competence), coordinating it with new evidence by means of strategies that are inferentially sound and applied in a consistent manner (metacognitive competence). In total absence of such competence, evidence and theory are not represented as distinct entities. In this case, new evidence may lead to modification of theory (as it does even among very young children), but the process takes place outside of the individual’s conscious control (Kuhn, 1989). (p. 12)

According to this position, one must be able to assess how the results of particular empirical tests provide evidence that bears on one’s current theories (see also, Demetriou & Raftopoulos, 1999; and Bickhard, 1999). Theories of case-based reasoning provide a similar account of representational change (see, e.g., Schank, Berman, & Macpherson, 1999).

In the second type of process, which we call redescription, information about the domain is extracted from the current representation as it is used. This information is then used to create a new, more sophisticated representation. In general, redescription involves forming a new representation based on activity with one’s current representation. Karmiloff-Smith’s (1992) work on representational redescriptions provides the broadest statement on the role of redescription in cognitive development. She proposed that children’s procedures begin as implicit and eventually come to be represented explicitly. As a procedure is mastered, information embedded in the procedure is extracted and used to represent the problem. These explicit representations can be manipulated and connected to previous knowledge. For example, in the domain of number, Karmiloff-Smith (1992, p. 104) suggested that children’s use of the counting procedure eventually results in the explicit representation: 1 + 1 = 2. This is a direct consequence of representing 1 and 1 as distinct entities and then applying the conventional rules of the number system to create the new representation.
representation of information about number. The counting procedure has a considerable amount of knowledge embedded in it, but much of this knowledge (e.g., ordinality) is initially unavailable to the child. As the counting procedure is practiced and mastered, the information embedded in the procedure becomes available as part of an explicit representation of number. For example, children begin to understand that 4 not only follows 3, but also that it indicates a larger set (see also Case & Okamoto, 1996). The application of the new representation to a problem (e.g., deciding whether to ask for 3 or 4 scoops of ice cream) requires metacognitive processes as well. Representational redescription is a process that builds the cognitively accessible representations upon which metacognition can operate (see Karmiloff-Smith, 1992, p. 16).

More recently, a number of researchers have proposed process accounts of how redescription might occur. For example, Kotovsky and Gentner (1996) proposed that redescription occurs as part of the progressive alignment of structural relations. They demonstrated that concentrated experience with comparisons facilitates the extraction of higher order relations from one’s current representation. Children whose comparisons were based on embedded relational information were better able to use that relation across domains. For example, the higher order relation of symmetry was extracted from the object-size and object-saturation domains and used to make relational comparisons across domains. Kotovsky and Gentner suggested that redescription may occur very locally and that the accumulation of these local changes may result in larger insights.

Siegler, Shrager, and Crowley proposed that redescription occurs through the operation of metacognitive processes on stored information about the effectiveness of each strategy (Crowley, Shrager, & Siegler, 1997; Shrager & Siegler, 1998; Siegler, 1996). This hypothesis, which is part of a larger theory of adaptive strategy choice, is instantiated in a computational model called the strategy choice and discovery simulation (Shrager & Siegler, 1998). The SCADS model posits that children’s problem solving is governed by two processes, an associative process and a metacognitive process. The associative process relies on a database about the strategies in a child’s repertoire. The system records the results of the entire strategy in terms of speed and accuracy, the partial results of each operation within the strategy, and the order of the operators. The associative process uses this database to adaptively select strategies that are efficient and accurate.

As associations between strategies and problems become increasingly strong, the metacognitive system begins to allocate attentional resources to strategy change. The metacognitive system applies strategy-change heuristics to existing strategies, drawing information from the current contents of working memory (i.e., information about the current operators) and the database described above. One heuristic attempts to identify and eliminate redundant operators. A second heuristic maintains the ordering of operators for the new strategy if the previous strategy was historically more successful when its component parts were executed in that order. The result is a new combination of operators, that is, a new potential strategy. The metacognitive system checks this new combination of operators for consistency with a set of predefined and constant criteria called a “goal sketch.” New combinations that are consistent with the goal sketch are applied as new strategies. In this way, the metacognitive system creates new strategies through redescription of existing strategies.

Finally, Schwartz and Black (1996) proposed that mental representations of physical systems become increasingly abstract over the course of repeated problem solving. Features of the problem that are not essential to the solution are “faded.” That is, the representation of the problem initially contains considerable detail. As problems are repeatedly solved, aspects of the representation that are not attended to drop out of the representation. For example, Schwartz and Black asked college students to predict which way the final gear in an interconnected set of gears would turn, given the direction of a driving gear. They proposed that aspects of the system that are not important for solving the problem (e.g., the texture of the gears) are faded with repeated problem solving. Schwartz and Black also proposed that important features of the mental representation (e.g., the direction of the driving gear) are codified into verbal labels. Codifying occurs both for aspects of the initial problem representation and intermediate solutions constructed during problem solving. Note that codifying necessarily involves applying prior knowledge to the problem. Schwartz and Black (1996) argued, “that fading and codifying are natural outcomes of repetitious modeling” (p. 483). The joint action of these processes redescribes the initial representation.

Theory revision and redescription provide very different accounts of how representational change occurs. That is, these systems are not simple recastings of one another or alternate descriptions of the same underlying processes. For example, theory revision produces representational change when one’s current representation produces errors, but redescription produces representational change when one’s current representation is highly accurate. Therefore, a representational change that was produced through theory revision cannot have been the result of redescription and vice versa. However, these systems are not mutually exclusive in the sense that demonstrating the existence of one system would be evidence against the existence of the other. It seems quite likely, given the strength of the evidence for each system, that different systems are used at different times or in different situations. One important possibility is that the different systems operate at different times during development and that sampling a broad age range would reveal developmental shifts from one representational-change system to another. A second important possibility is that both systems operate as mechanisms of representational change, but that the specific system utilized changes as competence develops. That is, children across a wide age range utilize both systems, but each has a particular niche during the course of knowledge acquisition. One purpose of the current article is to investigate the contribution of both systems to representational change across a broad age range and across the acquisition of two levels of expertise in a problem domain. To study representational change across a broad age range, we selected a problem domain, gear systems, with which all age groups

---

1 Karmiloff-Smith (1992) proposed that knowledge proceeds through a hierarchy of representational levels. The levels differ in the degree to which the representations are available to the cognitive system. Movement through the hierarchy occurs through the process of representational redescription. As Karmiloff-Smith noted, the appropriateness of the hierarchical model and the process of representational redescription are separate issues (p. 23). The current article focuses on the process of representational change. To distinguish between Karmiloff-Smith’s theory of representational redescription and the more general process of redescription, we use the term representational redescription only in reference to her theory.
were familiar but had little problem-solving experience. A second purpose of the article is to provide additional data on the nature of the redescription process.

The Gear System Domain

*Initial Representations*

Past research has shown that children (Lehrer & Schauble, 1998; Metz, 1985) and adults (Perry & Elder, 1997; Schwartz & Black, 1996) have poor initial representations of gear systems; they initially understand only that the movement of each gear affects adjacent, interlocking gears. For example, Lehrer and Schauble (1998) found that second- and fifth-grade students, given a display like that in the upper left panel of Figure 1, understood that if one gear turned, the adjacent gear would also turn. However, many students were unsure whether the adjacent gear turned in the same or opposite direction. Past research has also shown that children and adults acquire knowledge about gear systems relatively quickly and that their representations of gear systems undergo multiple changes.

*The Acquisition of Gear System Representations and Strategies*

Verbal protocol, response time, and gesture data suggest that, for the vast majority of children and adults, the first appropriate representation of the gear-system problem is in terms of physical forces (Lehrer & Schauble, 1998; Schwartz & Black, 1996). This

---

**Figure 1.** Examples of five different types of gear systems. In each system the driving gear provides the force. This gear has a single arrow on it. The fuel gear has a shelf that holds a small pile of coal. A variable number of gears, labeled with letters, connect the driving gear to the fuel gear. Two chutes appear near the bottom of the fuel gear; a button labeled *Jams* appears next to it. Participants selected one of the chutes or the jams button to indicate their response. Open systems are in the top row. These systems have a single pathway between the driving gear and the fuel gear. The remaining systems have two pathways between the driving and fuel gears. The bottom gear system also contains an extraneous gear, labeled *H*, that is not relevant to determining the movement of the fuel gear.
representation supports a strategy in which the turning force of the driving gear is traced as it moves through the system. That is, participants literally simulate how the turning of the driving gear will cause an adjacent gear to turn, and then how the turning of that gear will affect the next gear in the sequence, and so on, until they determine the movement of the target gear. These simulations of the movement of the system are often accompanied by both gestures that trace the movement of each gear and verbal descriptions of the movements. We refer to this as the Figure 8 (F8) strategy, because people’s tracing of the movement resembles a figure eight. This approach, although somewhat inefficient, yields correct answers if done appropriately.

Evidence suggests that adults and children move to more advanced representations quite rapidly (Lehrer & Schauble, 1998; Schwartz & Black, 1996). In one of these advanced representations, the gear system is represented as an alternating sequence. This representation supports a strategy in which each gear in the system is categorized in alternation (i.e., clockwise, counterclockwise). For example, if the driving gear turns clockwise, the adjacent gear is categorized as counterclockwise, the next gear as clockwise, and so on. Importantly, the categorization is done without simulating the movement of any of the gears (see Schwartz & Black, 1996). Rather, the alternating nature of the gear system is used to determine how each gear, and ultimately the target gear, will move. For example, consider the performance of a child in Lehrer and Schauble’s (1998) study. The child was asked to predict which way the final gear in a three-gear chain would turn, given the direction of the first gear. The child responded correctly and offered the following explanation:

Because there is one gear in between. This gear (blue) goes this way (clockwise), which makes this gear (red) go the opposite way, so the green gear goes the opposite way of the red gear. It’s the opposite of the one before it. So if there is one in between them, it just goes the same. (p. 17)

Lehrer and Schauble found that 25% of second-grade and 50% of fifth-grade children used this approach on some problems. We refer to this as the left–right (LR) strategy, because children tend to classify the gears as left or right turning (as opposed to the more appropriate, but more difficult, clockwise designations). This approach is considerably more efficient than F8 and yields correct answers when done appropriately. The alternating sequence representation may also support a more advanced strategy in which every other gear is categorized (i.e., left, skip one gear, left...), thereby eliminating the need to label each gear. This modified approach, which we call the skipping strategy, is slightly more efficient.

Past research has shown that many adults also eventually represent the problem mathematically. The parity of the number of gears in the problem (i.e., odd or even) determines whether the target will turn in the same direction as the driving gear (assuming that all gears are relevant). Some adults count (or subitize) the number of gears and identify the number as odd or even; this yields information about the direction of the target gear. (If two sets of gears connect the driving and target gear, as in the middle panels of Figure 1, the parity of each set must be the same or the system will jam). We refer to this as the counting parity strategy, because participants first count the number of gears and then use parity information. This approach is extremely efficient and accurate.

Consistent with previous research (e.g., Alibali, 1999; Lemaire & Reder, 1999; Siegler & Jenkins, 1989), we assume that as people construct more advanced representations they retain the strategies generated by their prior representations. For example, as children’s representations of the multiplication facts become increasingly strong, their use of retrieval as a solution strategy increases. However, their previous strategies that were developed prior to learning the multiplication facts, such as repeated addition, are also retained and used when necessary (Lemaire & Siegler, 1995).

Predictions: Effects of Accuracy and Response Time

The theory revision and redescription processes make distinct predictions about the precursors of discovery. We first focus on predictions about the effects of accuracy and response time. Further predictions about the effects of specific strategies on discovering new, more sophisticated strategies are presented in the Results section.

Theory Revision

In the theory revision process, new representations are generated when the current representation produces errors that are identified as such by the metacognitive system. Therefore, a history of low accuracy should predict discovery, and discovery should be directly preceded by longer response times indicative of metacognitive reflection.

Redescription

The accounts of the redescription process discussed above make common predictions about how accuracy and response time will be related to discovery. Discovery will be predicted by high accuracy as mastery is achieved with the current representation. Mastery allows the important relations to be extracted either through the application of metacognitive heuristics as attentional resources become available (as in SCADS) or through fundamental processes such as structural alignment (Kotovsky & Gentner, 1996) or fading and codifying (Schwartz & Black, 1996). For example, Schwartz and Black (1996) showed that college students discovered the LR strategy by extracting information embedded in the F8 strategy. That is, the action of tracing the alternating motion of the gears (i.e., the F8 strategy) provided crucial information about the alternating nature of the system itself; information that forms the basis for the LR strategy. Applying a new representation to the problem may also require metacognitive processes. Therefore, discovery should be preceded by longer response times as the metacognitive system engages (Crowley et al., 1997).

Summary

The theory revision and redescription processes both predict that response times will increase prior to discovery. Theory revision predicts that a history of low accuracy will precede discovery; redescription predicts that a history of high accuracy will precede discovery. An additional consideration was whether participants had used a higher strategy prior to discovering the strategy of interest. These participants may be falling back to the strategy of interest, rather than discovering it in the usual sense. Previous research suggests that children and adults may cycle through discovering new strategies and falling back on previous effective
strategies. As mentioned above, further predictions for each process are presented with reference to specific strategies in the Results section.

Method

Overview of the Task

Gear system problems were presented in the context of a game-like, computerized train race. Participants were told that they were the conductor of a train and that their train was competing against an opponent train controlled by the computer. The trains had to stop at fueling stations along the race route to obtain fuel. Each fueling station was a gear system. The fuel was positioned on a target gear that could drop the fuel down one of two chutes on either side of the target gear. The participant had to predict the movement of the target gear to obtain the fuel. A driving gear in each system provided the initial force. Participants were asked to select the appropriate chute or, if they thought that the gear system would not work (i.e., literally jam because of opposing forces), to select a button labeled Jams. Feedback about each decision was provided.

Each participant completed two sessions. Two sessions were used because of time constraints with the school age children and because we believed children’s attention to the task would wane if a single session lasted more than an hour. In the first session, participants received one of four different types of problems: small open, small closed, large open, or large closed. Small systems had four or five gears, large systems had seven or eight. Open systems had one path of gears between the driving and target gears. Closed systems had two paths of gears between the driving and target gears. Examples of the different types of gear systems can be seen in Figure 1. The second session consisted of the same set of problems for all participants. These problems included all the combinations of size (i.e., small, large) and structure (i.e., open, closed), as well as systems with an additional, extraneous gear that was not relevant to the solution. Systems with extraneous gears were included as one way to address the generalizability of children’s representations, an issue that we do not address in the current article.

Participants

Forty-three third-grade, 29 sixth-grade, and 56 college students participated in the study. Data from an additional 4 third-grade and 4 college students were not included in the analyses because of computer errors during the experiment. College students participated as one option to earn extra credit for a course. Third- and sixth-grade students were recruited from local public and private schools and from an after-school program and participated as volunteers.

We recruited a larger sample of third graders than sixth graders because the base-rate probability of discovering new strategies is lower for third graders. Similarly, we recruited a larger sample of college students because the probability of these older participants discovering a strategy of interest on the first trial is much greater. As we describe in the results, participants who discover a strategy on the first trial cannot be included in the analyses, because they have no history from which predictors can be derived.

Materials and Design

Participants solved gear problems as part of a computerized train race. Gear problems were presented as fueling stations in the race. For each problem, participants were asked to judge which way a target “fuel” gear in the system would turn. If their judgment was correct, their train would either “catch” the fuel or avoid waiting for a system that jammed. The train race was created using Supercard 4.0 (Allegiant Technologies, 1996). Participants used the mouse to indicate which way the gear should move. Each gear system was composed of a green driving gear, a variable number of connecting blue gears, and a red fuel gear. The fuel gear in each system had a small shelf protruding from it that held a pile of coal. The participant’s task was to predict which way the fuel gear would turn. The driving gear had an arrow showing that it turned to the right (i.e., clockwise). To facilitate verbal description of the gear system, each connecting blue gear was labeled with a letter. Each of these gears also displayed two blue arrows that pointed in opposite directions. Participants could click on either arrow to indicate which direction they thought the gear would turn. Once selected, the arrow turned red and the opposite arrow faded. Participants could set these arrows in any order (or not at all) and change them as many times as they wished.

Two chutes were located under the fuel gear. The chutes were positioned such that, if the fuel gear turned clockwise, the fuel would slide off the chute on the right; if it turned counterclockwise, the fuel would slide off the left chute. Participants indicated the direction the fuel gear turned by clicking on the corresponding chute; clicking on a chute highlighted it. In addition, a button labeled Jams was located next to the fuel gear. If a participant thought the station would jam, he or she would select this button. Once a chute or the jams button was clicked, the participant’s train appeared in the lower portion of the screen in one of three positions. If a chute was selected, the train moved underneath that chute in order to catch the coal. If the jams button was selected, the train moved to the far lower-left-hand side of the screen to give the appearance of preparing to leave the station.

Participants could change their answer (i.e., the selected chute or the jams button) as many times as they wished. When the participant was satisfied with his or her solution, he or she clicked on the train and moved on to the next screen. The computer recorded all the participant’s selections and final response.

Participants completed two races, an initial race and a standard race. For the initial race, participants were randomly assigned to one of four conditions formed by factorially combining the two levels of size and structure (i.e., small–open, small–closed, large–open, large–closed). Small systems had four or five gears, large systems had seven or eight. Open systems had a single pathway of gears from the driving to the fuel gear. Closed systems had two pathways. The problems within each condition were presented in random order.

The standard trials consisted of 32 stations comprised of all possible combinations of gear system size (small, large), structure (open, closed), and extraneous gear (present, absent). There were four trials within each of these eight conditions. Two trials contained the smaller number of gears for that level of system size (i.e., four for the small size, seven for the large size) and two contained the larger the number of gears (i.e., five for the small size, eight for the large size). The problems were presented in random order.

Four practice trials preceded the start of both the initial and standard races. These trials consisted of two-three-gear problems and two-six-gear problems. One system of each size was open and the other was closed. The practice trials were presented as a very short separate race and were used to familiarize the participants with the program. The practice trials were identical in all details (e.g., feedback sounds and screens, use of a map, etc.) to the substantive trials.

Once participants made their final decision about a station, an opaque screen appeared that covered the driving and connecting gears but not the fuel gear. For nonjamming systems, participants saw the fuel gear turn and drop the fuel. For jamming systems, participants saw pieces of the connecting gears fall from behind the opaque screen (indicating that the system had jammed). If the participant correctly predicted the outcome, the fuel gear either landed in his or her train (nonjamming systems) or the train sped out of the station (jamming systems). Next, a screen appeared showing the two trains in motion. If the participant had given the correct answer, his or her train moved faster than the opponent train. Whether the participant’s train was gaining or losing speed was also stated explicitly on the screen (e.g., “You’re gaining speed!”). These feedback screens included appropriate sounds (e.g., a trumpet sounding for catching the fuel, slow chugging for losing speed, etc.) to accompany the visual feedback. Finally, a map was
displayed with station markers and representations of the two trains; the trains moved on the map after each fueling station. Correct answers resulted in more movement of the participant’s train on the map relative to the opponent train.

**Procedure**

Participants were tested individually and were told that they would be playing a game in which they were a train conductor in a train race. Their goal was to beat another train to the finish line; the speed of their train depended on obtaining fuel. They were told they would have to stop at fueling stations along the way and decide whether they could get fuel there. The experimenter then selected a button on the computer screen to initiate the start of the race. At the first screen, a computer-generated voice invited participants to engage in a practice race. Participants saw a map of the race course with station markers on the right side of the screen and were shown their blue train and their opponent’s red train on the map. During the race, they were able to see this map to determine who was winning.

At the next screen, participants were shown the outside of the fueling station. They were told to click on a button labeled Fuel Up to see the inside of the station. The inside of the fueling station contained the gear system. Participants were told that the green gear was the driving gear that made the entire system move and that the red arrow on this gear indicated that it turned to the right (i.e., clockwise). The experimenter explained that the red gear was the fuel gear and pointed out the coal sitting on the shelf attached to this gear. Participants were told that they would be asked to predict whether the fuel gear would turn right (dropping the coal to the right) or left (dropping the coal to the left) or whether the whole system would jam. The experimenter demonstrated how participants would indicate their decision by clicking on the chutes or the jams button, and how the train would be positioned according to the choice that was made.

Participants were told that some people found it helpful to look at the blue gears in the system and determine which way they turned in order to decide what the fuel gear would do. The experimenter demonstrated how clicking on the arrows on the blue gears worked. Each blue gear was labeled with a different letter so that it could be easily identified. Participants were asked to think aloud when making their decision about the fuel gear and were told that if they talked about one of the blue gears, they should refer to it by its letter name.

When participants made a final decision about what the fuel gear would do, they were asked to click on the train that appeared on the screen in response to clicking the chute or jams button. The experimenter then described the feedback screens. Participants were told that a correct decision would help him or her gain time and that an incorrect decision would cause them to lose time in the race. The experiment took place over two sessions. The initial race was conducted on the first day and lasted approximately 30 min. The second race was conducted between 24 and 48 hr later and lasted approximately 45 min.

**Strategy Coding**

Participants were asked to think aloud as they solved each problem. The experimenter coded each participant’s strategy on each trial. Coding was based on the participant’s verbalizations and actions during his or her problem solving. If the participant did not think aloud during a particular trial, the experimenter asked for an explanation of how he or she had solved the problem before going on. The entire session was videotaped to allow for reliability coding.

On the basis of past research (Metz, 1985; Perry & Elder, 1997; Schwartz & Black, 1996) and a pilot study, we coded each trial for four major strategies: F8, LR, skipping, and counting parity. The F8 strategy is characterized by tracing the pattern of movement from the driving to the target gear. The LR strategy is characterized by categorizing the gears using an alternating sequence (i.e., left, right, left, right, etc.). The skipping strategy consists of categorizing every other gear as turning in the same direction (i.e., left, skip adjacent gear, left, etc.). With the counting parity strategy, participants determine the direction the fuel gear will turn by counting the number of gears in the system and noting whether the total is odd or even. If the system has an odd number of gears, the target gear turns in the same direction as the driving gear; if the system has an even number of gears, the target gear turns in the opposite direction (assuming that only relevant gears are counted).

We also coded whether participants recalled the answer from an identical previous problem (recall), guessed the answer (guessing), or used one of two incorrect strategies (other). The two “other” strategies were (a) referring to the prior problem (e.g., “the fuel gear turned left last time, so this time it will turn right”) and (b) asserting that the direction a gear turned was determined by its color (e.g., “the blue gears turn left, so the green gear will turn right”). The strategies are summarized in Table 1. Three raters independently coded the entire session of five randomly selected cases (approximately 6 hr of videotaped trials). The raters showed perfect agreement on 96% of trials.

**Results**

**Overview**

First, we describe students’ strategy use and how strategy use changed across trials for each age group. Next, we present data on

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Strategies Coded for the Gear System Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strategy</td>
<td>Description</td>
</tr>
<tr>
<td>Nonclassifiable</td>
<td>Offered an explanation that was not logically coherent.</td>
</tr>
<tr>
<td>Guessing</td>
<td>Reported guessing the answer.</td>
</tr>
<tr>
<td>Other</td>
<td>Either (a) asserted that the answer to the current problem depended on what happened in the previous problem or (b) asserted that gears of the same color turned the same direction.</td>
</tr>
<tr>
<td>Figure 8</td>
<td>Traced the pattern of movement from the driving gear to the fuel gear.</td>
</tr>
<tr>
<td>Left-right</td>
<td>Categorized the gears using an alternating sequence.</td>
</tr>
<tr>
<td>Skipping</td>
<td>Categorized every other gear as turning in the same direction.</td>
</tr>
<tr>
<td>Counting parity</td>
<td>Counted the number of gears and noted that, because the number was odd or even, the fuel gear had to turn in a particular direction.</td>
</tr>
<tr>
<td>Recall</td>
<td>Reported recalling the answer from a previous problem.</td>
</tr>
</tbody>
</table>

**Note.** The names of the coded strategies are on the left. The descriptions of the participant’s verbalizations, actions, or both are on the right.
the accuracy and response time for the different strategies. We then consider students’ discoveries and test the predictions of the theory revision and redescription processes for the discovery of the two most prevalent strategies. All reported effects are significant at the .05 level.

**Strategy Use and Efficacy**

Figure 2 shows the percentage of strategy use across trials for each grade. The results for the third graders are in the top panel. As can be seen in the figure, a substantial proportion of third-grade students guessed or used one of the “other” strategies (i.e., a strategy not based on the physical system). However, there was a considerable proportion of F8 use across trials and a smaller proportion of LR use. As the figure shows, there was very little change in distribution of strategies across trials, Friedman two-way analysis of variance, \( \chi^2(43, N = 43) = 19.63, \text{ ns.} \)

The middle panel of Figure 2 shows the results for the sixth graders. The majority of sixth graders began solving the gear system problems with the F8 strategy, although a substantial proportion used an “other” strategy. Over trials, the LR strategy became increasingly prevalent, Cochran’s \( Q(43) = 167.48, \) and use of the F8 strategy declined, Cochran’s \( Q(43) = 77.79. \) The counting parity strategy was seldom used; none of the sixth-grade students ever used it regularly.

The lower panel of Figure 2 shows the proportion of strategy use for the college students. Approximately 70% of the college students first solved the problems with F8; 20% used LR first. As the figure clearly shows, the use of LR increased in the second session, Cochran’s \( Q(43) = 155.21, \) and the use of F8 dropped off precipitously, Cochran’s \( Q(43) = 406.75. \) Also, the college students’ use of counting parity increased substantially in the second session, Cochran’s \( Q(43) = 249.32. \)

Figure 3 shows the proportion of correct responses (left panels) and mean response time (right panels) for each strategy by grade. As would be expected, guessing resulted in low accuracy for all grades. Similarly, the use of an “other” strategy tended to be very inaccurate for the third and sixth graders, but less so for the college students. The “other” strategies also had relatively short response times.

The third graders were reasonably accurate with the F8 strategy, although the LR strategy was both more accurate and faster. Surprisingly, third graders were not faster or more accurate with the skipping strategy than with F8 or LR. The sixth-grade and college students were highly accurate with the F8, LR, and counting-parity strategies. However, the LR strategy was faster than the F8 strategy for both age groups. For the college students, the skipping and counting parity strategies were faster than LR.

In summary, the older students (i.e., sixth grade and college) increasingly used more effective strategies over time, both in terms of speed and accuracy. At the group level, change in the third graders’ performance was much more modest. However, over the course of the experiment, students from all three age groups displayed new strategies.

**Discovery of New Problem Representations**

Because past research suggests that different problem representations are necessary for their construction, the first uses of the F8, LR, and counting parity strategies were considered to indicate the discovery of a new problem representation. For simplicity of exposition, we refer to these first uses as discoveries of strategies. Discoveries that occurred after the first trial of a session (i.e., Trials 2–12 in Session 1, Trials 3–32 in Session 2) were included in the analyses. Students who used the strategy on the first trial may well have discovered it on that trial, but we have no history of their performance on which to base predictions. Therefore, these first uses are not analyzed. These participants are usually referred to as left censored in time-series nomenclature.

The F8 and LR strategies were discovered quite frequently across age groups. The F8 strategy was discovered by 11% of third graders, 34% of sixth graders, and 16%, of college students. The LR strategy was discovered by 14% of third graders, 38% of sixth graders, and 54% of college students. Only a single third and a single sixth grader discovered counting parity, although 45% of college students did.

Because discoveries of F8 and LR strategies were reasonably distributed across age groups, and because they reflect very different problem representations, we focus on the discovery of these strategies. We do not report analyses regarding the counting parity strategy, because it was discovered almost exclusively by college students.

**Examining the Precursors of Discovery**

*Event history analysis.* A class of statistical techniques has been developed to handle the unique problems of longitudinal event data (see Allison, 1984, for an accessible introduction). An event is a discrete occurrence, such as a birth, job change, correct recall of a test item, or discovery of a new problem representation. This class of techniques, called event history analysis, provides a system for analyzing categorical, longitudinal data in a familiar, regression-like format. It also allows for time-varying predictor variables and appropriately handles right-censored cases, issues that are problematic when standard ordinary least squares (OLS) multiple-regression techniques are applied to longitudinal event data. Time-varying predictor variables are a common feature of longitudinal data, regardless of whether the period of study is multiple years or multiple trials, as in the current microgenetic design. For example, in the current study, we are interested in how the accumulation of errors affects the probability of discovering a new strategy. Of course, the number of errors a child has made changes from trial to trial, depending on his or her performance and, therefore, this predictor variable is time-varying. (When time is tied to a particular measurement interval or to trials, as in the current study, time can be treated discretely.)

Right censoring is another common characteristic of longitudinal event data. A case is right censored when the study ends before the participant shows the event of interest. For example, a child who never discovers (or uses) the F8 strategy would be right censored with regard to F8. When standard OLS techniques are applied to longitudinal data, these cases are usually excluded from the sample, a practice that can substantially affect the results.

Although a comprehensive account of event history analysis is beyond the scope of the current discussion, a brief description of the general approach will show that it is an easily understandable method with many similarities to more familiar techniques. Two concepts are central to understanding event history analysis. The first is the risk set. The risk set consists of the participants who have not yet experienced the event. For example, in the current
Figure 2. Percentage of strategy use for participants in each grade over trials. The smaller panels on the left show the distribution of strategy use across the initial 12-trial session. The larger panels on the right show the distribution across the 32 standard trials.
study, all the participants who did not use the F8 strategy on Trial 1 are in the risk set on Trial 2, regardless of whether they discover F8 later in the experiment. Obviously, the risk set changes across trials; participants who discover the strategy are no longer at risk for the event and, therefore, drop out of the set. The second central concept is the hazard rate. The hazard rate is simply the probability of the event occurring on any given trial. For example, the hazard rate for discovering F8 on Trial 2 can be estimated by dividing the number of F8 discoveries on Trial 2 by the number of participants in the risk set for that trial. Therefore, the hazard rate is the population parameter of interest. In the current context, the hazard rate reflects the probability of discovering a strategy. Because the dependent variable is dichotomous (i.e., discovery or no discovery), we estimated explanatory models using logistic regression.

Analytic approach and predictor variables. In the current study, we used discrete-time event history analysis to test predictions of the theory revision and redescription processes. To limit the number of models tested for each type of discovery, we capitalized on the opposing predictions regarding the effect of accuracy. The theory revision process predicts that low accuracy should be associated with discovery, whereas the redescription process predicts that high accuracy should be associated with discovery. Therefore, for each type of discovery, we first analyzed the effects of accuracy and response time. A significant effect of accuracy would disconfirm one process as the explanation for that discovery. We then assessed whether these effects depended on grade, and conducted additional analyses to evaluate more specific predictions of the remaining process.

Because we were analyzing how the probability of discovering a strategy changes across trials, an important issue was how the predictor variables were constructed from each participant’s history prior to the focal trial (where the focal trial is the trial currently under analysis). We initially constructed two different types of predictor variables for accuracy and response time. One type of predictor variable reflected each participant’s accuracy and response time across all trials prior to the focal trial. For example, if the current focal trial were Trial 10, the participant’s mean accuracy and mean response time prior to that trial (i.e., from Trials 1 to 9) would be used as predictors. For focal trials in the second session, these variables were computed across all previous trials, including those in the first session. The second type of predictor variable reflected each participant’s accuracy and response time across recent trials. These variables were constructed by taking the mean accuracy and response time of the previous trials. (For Focal Trials 2–5, these variables reflect the mean accuracy and response time of the previous trials.) Because these variables were intended to capture temporally recent effects, they were computed within session. That is, for trials in Session 2, these variables did not carry information from Session 1. Preliminary analyses showed that recent accuracy and response time were associated with strategy discovery, but mean accuracy and response time over the entire history were not. Therefore, except

<table>
<thead>
<tr>
<th>Grade</th>
<th>Mean Proportion Correct</th>
<th>Mean Response Time (secs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3rd</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6th</td>
<td></td>
<td></td>
</tr>
<tr>
<td>College</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3. Left: Mean proportion of correct responses for each strategy. Right: Mean response times for each strategy. The top, middle, and bottom panels show the results for third-grade, sixth-grade, and college students, respectively. NC = nonclassifiable strategies; F8 = Figure 8 strategy; LR = left–right strategy; CP = counting parity. Strategies with less than three uses within a grade are not shown.
where otherwise noted, we used the “recent” variables in the substantive analyses.²

**Discovery of F8.** We tested the effects of accuracy and response time on the probability of discovering F8, also including whether participants had previously used a higher strategy (i.e., LR, skipping, or counting parity). Five predictor variables were entered into the equation: grade, recent accuracy (RA), recent response time (RRT), prior use of a higher strategy (PRIOR), and the interaction between prior use of a higher strategy and recent accuracy. Grade and RRT contributed significantly to the model, change in log LR $\chi^2(1) = 5.85$ and 18.31, respectively, $B_s = .17$ and .03, respectively. The effects of RA and PRIOR were not significant, change in log LR $\chi^2(1) = 0.08$ and 3.82, respectively. However, the RA × PRIOR interaction did contribute significantly to the model, change in log LR $\chi^2(1) = 4.46$.

The effect of grade is straightforward: older children were more likely to discover F8, even with the other predictor variables in the model. When the interactions between grade and RA and grade and RRT were added to the model, there was no significant improvement in fit, change in log LR $\chi^2(1) = 0.89$. This makes considerable sense given that the participants in the risk set against which these discoveries were compared had not yet used F8 or any other appropriate strategy and, therefore, had very low accuracy rates (mean proportion correct = .37). The model predictions for participants who had used a higher strategy prior to discovery are on the left. Predictions for participants who had used a higher strategy are on the right.

Two In a separate set of analyses, we examined whether (a) the size and structure (i.e., open vs. closed) of the gear systems in the first session and (b) the individual’s history of exposure to different types of gear systems (i.e., size, structure, extraneous gears) across both sessions affected the probability of discovering F8 and LR. Neither the size nor the structure of the gear systems in the first session significantly affected the probability of discovering F8 or LR, largest change in log LR $\chi^2(1) = 0.96$. Similarly, the history of exposure to different types of gear systems did not predict discovering F8 or LR, largest change in log LR $\chi^2(1) = 2.02$.

³ Note that the model predicts the probability of an individual discovering the strategy on a single focal trial. Therefore, small probability values are expected. For example, assuming that discoveries are randomly distributed across trials and that all participants discovered the strategy, the average probability of discovery per trial would be .10. When discoveries occur later in the course of the study, the probability decreases dramatically. For example, again assuming that all participants discovered the strategy, but that discoveries were limited to the second session, the average probability of discovery per trial is .04.

The probability of discovery becomes even smaller, if, as in the current study, nondiscoverers are present. These participants effectively remain in the denominator, thus reducing the probability of discovery further.
accuracy had a substantial effect on the probability of discovering F8, change in log LR $\chi^2(1) = 11.26$.

It is worth noting that the individual data patterns are quite consistent with the results from the model. For example, 83% of participants who discovered F8 and had not used a higher strategy prior had RA of 50% or less. Similarly, these same participants had long RRTs on the trials immediately preceding discovery; the RRTs of 67% of these participants were more than twice the average RRT of participants remaining in the risk set ($M = 6.76$ s). Only a single participant who discovered F8 had an RRT that was less than this mean on their discovery trial.

These results provide strong evidence against the redescription process as the source of discovering the F8 strategy. The redescription process predicts that high accuracy, indicative of mastery, should precede discovery. This prediction is in direct conflict with the negative relationship between accuracy and discovery observed for participants who had previously used a higher strategy. Similarly, the low accuracy rate for participants who had not used a higher strategy is inconsistent with this prediction.

However, the theory revision process fits the observed data quite well. Participants who had not used a higher strategy prior to discovering F8 had not yet developed a way to appropriately represent the physics of the gear system. These students made considerable errors right up to their discovery of F8 and had long response times just prior to the discovery. For participants who used a higher strategy prior to their first use of F8, low accuracy and long response times predicted discovery. These participants may have been discovering F8 through theory revision. Alternatively, they may have already had F8 in their repertoire and had fallen back on the F8 strategy when their higher strategy became inefficient.

**Further tests of the theory revision process: No higher prior strategy.** To further test the theory revision process for the participants who had not used a higher strategy prior to discovering F8, we examined the effects of guessing and of proposing an incorrect strategy. According to the theory revision position, participants must propose some theory about the causal relationships within the problem in order to change that representation in response to contrary evidence. Therefore, participants who guess rather than propose a causal explanation should be unlikely to discover a new strategy. Participants who propose an incorrect causal relationship (i.e., “other” strategies) should be likely to make such a discovery. Note that the guessing and other strategies are nearly identical in terms of their effectiveness (i.e., the mean accuracy for guessing and the mean accuracy for “other” were identical for participants in the risk set, $M = 0.50$; mean response times were nearly identical, $M = 14.89$ s for “other,” and $M = 14.65$ s for guessing).

To test this prediction, we calculated the proportion of use for the guessing and “other” strategies across the five previous trials. As predicted by the theory revision process, when these variables were added to the model they had opposite effects. The proportion of recent guessing was negatively associated with discovering F8 ($B = -3.13$), change in log LR $\chi^2(1) = 11.25$, but the proportion of recent “other” use was positively associated with F8 discovery ($B = 1.82$), change in log LR $\chi^2(1) = 4.64$. Participants who proposed an incorrect solution were more likely to discover F8; participants who guessed were much less likely, despite the fact that both of these approaches produced chance-level performance. Neither the effect of using an “other” strategy nor the effect of guessing depended on grade, change in log LR $\chi^2(1) = 0.00$ and 0.54, respectively.

**Further tests of the theory revision process: Higher prior strategy.** For participants who had used a higher strategy prior to discovering F8, we were interested in whether the first use of F8 was the product of theory revision or of falling back on a strategy that was already in their repertoire. If their first use of F8 was actually the employment of a backup strategy, then we would expect their performance to be indistinguishable from that of participants who had demonstrably used F8 as a backup strategy (i.e., used F8 early in the experiment, used a higher strategy for a number of trials, and then used F8 again). Therefore, we compared the first use of F8 by participants who had previously used a higher strategy to the first reuse of F8 by participants who had used F8 on the first trial (and were, therefore, left censored in previous analyses) and used a higher strategy on the five trials preceding the focal trial.

The following variables were entered as predictors: grade, RA, RRT, whether the F8 strategy had been observed before (F8PRIOR), and the interactions between RA and F8PRIOR and RRT and F8PRIOR. Of particular interest was whether the effects of RA and RRT would depend on whether the F8 strategy had been previously observed. Neither interaction contributed significantly to the model, change in log LR $\chi^2(2) = 0.00$. RA and RRT appeared to have the same predictive effects for participants who demonstrably fell back on the F8 strategy and for those who used a higher strategy prior to their first use of F8.$^4$

**Discovery of LR.** Analogously to the analyses conducted for the discovery of F8, we first tested the effects of RA and RRT on the discovery of LR, including in the analysis whether participants had used a higher strategy (i.e., skipping or counting parity). Five predictor variables were entered: grade, RA, RRT, PRIOR, and the interaction between PRIOR and RA. Grade, RA, RRT, and PRIOR all had significant effects, change in log LR $\chi^2(1) = 14.71$, 4.14, 10.18, and 9.62, respectively. The effects of grade, RA, RRT, and PRIOR were all positive ($B = 0.16$, 0.68, 0.02, and 1.83, respectively). The RA $\times$ PRIOR interaction was also significant ($B = -1.71$), change in log LR $\chi^2(1) = 7.42$.

Older participants were more likely to discover LR. When the interactions between grade and RA and grade and RRT were entered into the equation, the fit of the model did not improve significantly, change in log LR $\chi^2(2) = 3.13$, $p > .05$. The effects of RA and RRT did not appear to depend on grade.

The top panel of Figure 5 shows the model predictions for RRT: longer RRT increases the probability of discovering LR. The nature of the interaction between PRIOR and RA is shown in the two lower panels of Figure 5. For participants who had not previously used a higher strategy (left panel), discovery of LR was significantly predicted by an increase in accuracy on recent trials, change in log LR $\chi^2(1) = 5.29$. The low probability of discovery for these participants is a consequence of their discovering LR relatively late in the study and the inclusion of nondiscoverers in the risk set (see footnote 3). For participants who had used a higher strategy prior (right panel) to discovering LR, a decrease in accuracy predicted discovery.

Again, note that the individual data patterns are consistent with the results from the model. For example, 79% of participants who

---

$^4$ We thank an anonymous reviewer for suggesting this analysis.
discovered LR but had not used a higher strategy had perfect accuracy on the recent trials. Similarly, compared to the mean for the rest of the risk set \((M = 9.52\) s), these participants also tended to have long RRTs; 74% had RRTs of 12 s or longer.\(^5\)

This pattern of results strongly suggests that different processes are responsible for discovering LR depending on the participant’s current representation of the problem. For participants who had not previously used a higher strategy, discovery of LR is predicted by high accuracy and longer response times. These effects are consistent with redescription. For participants who have previously used a higher strategy, discovery was predicted by low accuracy and long response times. These effects are consistent with either theory revision or falling back on a strategy already in their repertoire. However, consistent with the fall-back hypothesis, the predictive effects of accuracy and response time on discovering LR for this group and for participants who had demonstrably used LR as a backup strategy did not differ, change in log LR \(\Delta \log LR = 0.06\).

**Further tests of the redescription process.** For participants who had not used a higher strategy, discovery of LR was predicted by high accuracy and long response times. This pattern of results is inconsistent with the theory revision process, but quite consistent with redescription. Therefore, we conducted another set of analyses to test additional aspects of the redescription process.

Past research with gear system problems has suggested that the discovery of LR is based on embedded information contained in the F8 strategy. Recall that Schwartz and Black (1996) presented evidence that the F8 strategy was replaced by the LR strategy through fading and codifying. Consistent with existing outlines of the redescription process, we assumed that the following processes occur during problem solving. Each time a child solves a gear system problem, a large amount of stored semantic information is activated. Some of this information is activated by encoding the problem itself (i.e., the color of the gears, whether the gears form an S or a U shape), but the child’s problem representation and actions also activate semantic information. Children using the F8 strategy will activate information about turning, because they represent the turning motions of the gears. Similarly, they may have information about alternating sequences activated because the direction of their gestures and labels for gears alternates (i.e., “this way, that way, this way, ...”). This alternating sequence relationship is the crucial piece of embedded information for discovery of the LR strategy.

Much of the information activated during problem solving is irrelevant, and activation of unattended and unrepeated information fades across trials. Semantic information that is repeatedly activated begins to be represented in long-term memory as part of the problem. For example, the fact that the target gear is red will become part of the problem representation stored in long-term memory. Similarly, information activated by children’s strategy use (e.g., turning and alternating sequences for the F8 strategy) will also become part of the representation. Finally, concentrated use of a strategy strongly activates semantic information in working memory. For example, use of the F8 strategy across multiple,

\(^5\) In general, it does not appear that the results from the model, here or elsewhere, were driven by outliers or unusual patterns of performance. For each set of analyses, we examined the residuals following the procedure recommended by Menard (1995). Elimination of candidate outliers did not substantially affect the pattern of results.
successive trials will make information about turning and alternating sequences highly salient in working memory.

Given these assumptions, we propose that redescription may occur in three different ways. First, redescription may depend on the quality of the representation in long-term memory. That is, a stable and complete representation of the alternating sequence information in long-term memory may allow a new problem representation to be constructed on the basis of that information.

Second, redescription may occur when the relevant information is highly activated in working memory during the course of problem solving. That is, regardless of the quality of the representation in long-term memory, when the alternating sequence information becomes highly activated in working memory it can be used as a basis for representing the problem.

Third, redescription may depend on both activation of the alternating sequence information in working memory and the quality of the representation of that information in long-term memory. For example, as a child accurately performs the F8 strategy, alternating sequences becomes part of the problem representation in long-term memory. If the child uses the F8 strategy over consecutive trials, alternating sequence information will become highly activated in working memory. To the degree that the activated information is already represented in long-term memory as part of the problem, it may be available for use as a basis for representing the problem.

If redescription depends on the quality of the long-term memory representation, then the degree to which an individual uses F8 accurately and repeatedly should predict discovery. As F8 is repeatedly and accurately used, the alternating sequence information embedded in the strategy becomes part of the representation in memory. Therefore, the interaction between total number of F8 uses and accuracy of F8 uses should predict discovery of LR.

If redescription depends on the activation of the alternating sequence information in working memory, then concentrated and accurate use of F8 should predict discovery of LR. That is, the interaction between the proportion of recent F8 use and the accuracy of recent F8 use should be associated with discovery of LR.

If redescription depends on both the quality of the representation in long-term memory and the degree of activation in working memory, then the effect of recent F8 use should depend on having a history of accurate F8 use (prior to and including recent trials).

We tested these hypotheses by fitting three additional models, one for each of the hypotheses above. These models tested the predictive effects of the following four variables (and interactions among them) on the discovery of LR: the frequency of F8 use on all trials preceding the focal trial (F8FREQ); the mean accuracy of trials on which F8 was used, computed across all trials preceding the focal trial (F8ACC); the proportion of recent F8 use over the five trials prior to the focal trial (RECF8); and mean accuracy of trials on which F8 was used, computed across the previous five trials (RECF8ACC).

In the model for the effect of the quality of the representation in long-term memory, we included the following predictor variables: grade, F8FREQ, F8ACC, mean response time on F8 trials (F8RT), the interaction between F8FREQ and F8ACC, and the interaction between F8FREQ and F8RT. The effect of F8FREQ was not significant, change in log LR $\chi^2(1) = 1.25, p = .26,$ and was in the wrong direction ($B = -0.09$). That is, increases in the use of the F8 strategy were associated with lower probability of discovering LR. The effect of accuracy was significant, change in log LR $\chi^2(1) = 7.21, B = 1.52,$ but response time was not, change in log LR $\chi^2(1) = 0.94$. F8FREQ did not interact significantly with accuracy or response time, change in log LR $\chi^2(1) = 0.01$ and 1.40, respectively. The model predictions for F8FREQ and F8ACC can be seen in the left panel of Figure 6.

The model for the effects of activation in working memory was constructed analogously, except that the RECF8 replaced the F8FREQ and RECF8ACC replaced F8ACC. Therefore, the predictor variables were grade, RECF8, RECF8ACC, F8RT, the interaction between RECF8 and RECF8ACC, and the interaction between RECF8 and F8RT. The effects of RECF8, RECF8ACC, and F8RT, and the RECF8 × RECF8ACC and RECF8 × F8RT interactions were all nonsignificant, change in log LR $\chi^2(1) = 0.38, 3.06, 2.32, 0.78,$ and 1.27, respectively. Because none of the predictor variables contributed significantly to the model (with the exception of grade), we did not plot the predictions.

**Figure 6.** Left: Model predictions for the long-term memory representation hypothesis. The probability of discovering the left–right (LR) strategy is plotted as a function of total Figure 8 (F8) use with a separate curve for different levels of F8 accuracy. Right: Model predictions for the hypothesis that activation in working memory and representation in long-term memory precede discovery. The probability of discovering LR is plotted as a function of the proportion of recent F8 use with a separate curve for different levels of F8 accuracy.
The model for the joint effects of activation and long-term memory representation was also constructed analogously. The predictor variables were grade, RECF8, F8ACC, F8RT, the interaction between RECF8 and F8ACC, and the interaction between RECF8 and F8RT. The effects of RECF8, F8ACC, and F8RT, and the RECF8 × F8RT interaction were nonsignificant, change in log LR \( \chi^2(1) = 0.41, 0.33, 2.38, \) and \( 1.22, \) respectively. However, the interaction between RECF8 and F8ACC was significant, change in log LR \( \chi^2(1) = 6.23. \) The nature of this interaction can be seen by examining the model predictions in the right panel of Figure 6. The predicted probability of discovering LR is plotted as a function of the proportion of recent F8 use (x-axis) with a separate curve for different levels of accuracy with the F8 strategy. The effect of RECF8 is slightly negative when accuracy is low, but it becomes increasingly positive at high levels of accuracy (i.e., approaches 1.0). This interaction did not depend on grade; when the RECF8 × F8ACC × Grade interaction was added to the equation, the fit of the model did not improve significantly, change in log LR \( \chi^2(1) = 0.00. \) This pattern of results is consistent with the hypothesis that both activation of the alternating sequence in working memory and the representation of that information in long-term memory are necessary for discovery.

It is worth noting that the interaction between RECF8 and F8ACC mediates the relationship between RA and discovering LR for participants who had not previously used a higher strategy. Adding RA to a model that includes the RECF8 × F8ACC interaction does not improve the fit of the model, change in log LR \( \chi^2(1) = 0.19. \) However, adding the RECF8 × F8ACC interaction to the model that includes the effect of RA contributes significantly to the fit of the model, change in log LR \( \chi^2(1) = 5.28. \) A history of accuracy with the F8 strategy in combination with recent, concentrated use of F8, not only predicts discovery of LR, but also explains the relationship between RA and discovery of LR.

Discussion

Our results show that discoveries in problem solving are not the result of a single system or set of processes for any of our age groups. Rather, it appears that, depending on their current representation of the problem, the processes underlying both children’s and young adults’ discoveries are very different. Participants whose representation of the gear system was inappropriate (e.g., predicted that all the blue gears turned the same way) appeared to be using theory revision to generate discoveries of the F8 strategy. These participants had very low accuracy prior to discovering F8. Longer response times also predicted discovery for these participants. Siegler and Jenkins (1989) reported a similar effect for children who discovered a new addition strategy; children’s response times became longer just prior to discovering the new strategy (see also, Alibali & Goldin-Meadow, 1993; Perry, Church, & Goldin-Meadow, 1988). We also found that using an incorrect strategy predicted later discovery, but guessing at the answer was actually negatively associated with discovery.

Participants who were already representing the gear system appropriately, as indicated by use of a higher strategy, appeared to be discovering F8 as a backup strategy. For these participants, first use of F8 was predicted by low accuracy with the higher strategies and longer response time. Furthermore, the predictive effects of RA and RRT were the same for these participants and for participants who had demonstrably used F8 as a backup strategy. These results are quite consistent with previous work demonstrating the use of backup strategies (see Lemaire & Siegler, 1995; Siegler & Shipley, 1995). When higher strategies fail, people fall back on less advanced strategies.

The LR strategy is based on a categorical representation of the gear system in which each gear is classified as left-turning or right-turning without actually representing the motion of the gears (Schwartz & Black, 1996). Participants who had not used a higher strategy (i.e., skipping or counting parity) appeared to discover LR through the redescription process. Discovery of the LR strategy was predicted by increased RA on recent trials and longer RRTs. We also found that concentrated use of the F8 strategy—a strategy that contains embedded information (i.e., alternating sequences) that is the basis for the LR strategy—in conjunction with a history of accurate F8 use predicted discovery of LR.

Participants who had used a higher strategy, either skipping or counting parity, prior to discovering LR showed a very different pattern of results. Consistent with the possibility that the discovery of LR for these participants was a backup strategy, first use was predicted by a decrease in RA and an increase in RRT. Note that this pattern of results directly parallels that for the F8 strategy. The fall-back hypothesis provides a parsimonious explanation of both F8 and LR discoveries for participants who had previously used a higher strategy.

Older students were more likely than younger students to discover new problem representations. However, we found no evidence for age differences in the underlying processes. Grade did not interact with RA or RRT, as would be predicted if there were a developmental shift in the use of the theory revision and redescription processes. Similarly, grade did not interact with the predictor variables that provide additional evidence about the theory revision process (i.e., recent use of guessing or other), nor did grade interact with the predictor variables that provided further evidence about redescription (i.e., F8ACC and RECF8).

Therefore, the present results suggest that, rather than changing developmentally, the processes responsible for representational change differ depending on the quality of the current representation. Given a problem representation and associated strategies that produce low accuracy rates, children and young adults use the theory revision process to discover new representations. Given a problem representation and strategies that accurately solve the problem, they extract embedded information and use that information as the basis for a new and improved representation. Finally, given a problem representation that generates a new high-level strategy, but a history of low accuracy and long response times with that strategy, children and young adults fall back on other strategies in the repertoire. All three processes appear to be involved in discovery but they occur at different points in the development of understanding and solving a problem. These processes seem to operate jointly in the development of understanding physical systems such as the gear system task. Whether these processes operate jointly in other domains is not currently known.

Our results also provide additional information about the details of the redescription process. We found that the quality of the representation of the alternating sequence information in long-term memory, as indexed by the interaction between the frequency and accuracy of F8 use across all prior trials, did not predict LR discoveries. Similarly, activation of the alternating-sequence relationship in working memory alone, as indexed by the interaction
between recent use of F8 and recent accuracy of the F8 strategy did not predict discovery of LR. However, the interaction between recent use of F8 and accuracy of F8 across all prior trials predicted discovery of LR. When F8 had been used with high accuracy in the past and was subsequently used in repeated succession, the probability of discovery increased. This is preliminary evidence in support of the hypothesis that representation of the relevant information in long-term memory and strong activation in working memory facilitate discovery of a new problem representation. We propose that as the F8 strategy is repeatedly and accurately used, the embedded alternating-sequence information becomes part of representation in long-term memory. Subsequent, concentrated use of the F8 strategy strongly activates this part of the problem representation in working memory and thereby creates a new way of representing the problem—as an alternating sequence. Implementing this new representation into a strategy that is appropriate for a particular problem may require additional, perhaps metacognitive, processes. Redescription creates new representations, but the application of these representations to the problem is not automatic. Therefore, discovering a new approach to a problem involves both building the appropriate representation and coordinating its application to the problem. A similar distinction is made in research on analogical problem solving; an analogy may provide access to a previous solution strategy but not facilitate mapping the strategy to the current problem (e.g., Ross & Kilbane, 1997).

Two of our findings regarding the theory revision process are noteworthy. First, we found that when accuracy was very low, longer response times on prior trials predicted discovery of a new representation. This is consistent with previous work cited above and with the idea that prior to a representational change the metacognitive system is engaged as part of the theory revision process. Second, we found that having recently proposed an incorrect hypothesis predicted discovery of a new representation. This finding creates something of a paradox when considered with previous work that demonstrated that children and adults have a strong tendency to confirm initial incorrect hypotheses (Klayman & Ha, 1987; Schauble, 1996; see also, Dixon & Tuccillo, 2001; Dunbar, 1993; Klahr, Fay, & Dunbar, 1993; Kuhn et al., 1995). Proposing an incorrect hypothesis is necessary to engage the theory revision process, but it also creates a tendency to confirm that hypothesis. These conflicting aspects of proposing an incorrect hypothesis may help explain why children often have such difficulty discovering new representations.

The analyses reported here were motivated by two prominent accounts of how representational change occurs. However, other factors beyond those considered here may also contribute to our understanding of these processes. For example, the SCADS model (Shrager & Siegler, 1998) includes the efficiency of each strategy and the operations within each strategy as important data for generating a new strategy. Clearly, efficiency and other important performance factors may drive representational change either through processes such as theory revision and redescription, or through some other set of processes yet unspecified. Chaining together the theory revision, redescription, and fall-back processes makes considerable sense for developing effective problem solving strategies. For example, when a child has an inappropriate representation of a problem, he or she will experience recurring errors and, therefore, search for a more effective representation. Once a moderately effective representation is in place, the patterns and regularities from the child’s repeated problem solving eventually coalesce to form alternative representations. The metacognitive system can select a potentially more effective representation from among these alternatives. However, previously effective strategies remain available and can be applied if the current strategy begins to fail or becomes too demanding. In this way, the theory revision process allows children and adults to construct some appropriate representation of the problem. The redescription process can then generate subsequent improvements on the basis of the activity that results from using that representation. The fall-back process provides the ability to return to previous representations in the face of new types of problems or new task demands. Jointly, these processes allow children and adults to discover increasingly appropriate problem representations.

References


Received May 21, 2001
Revision received May 28, 2002
Accepted May 29, 2002

---

**Call for Nominations**

The Publications and Communications (P&C) Board has opened nominations for the editorships of *Contemporary Psychology: APA Review of Books, Developmental Psychology*, and *Psychological Review* for the years 2005–2010. Robert J. Sternberg, PhD, James L. Dannemiller, PhD, and Walter Mischel, PhD, respectively, are the incumbent editors.

Candidates should be members of APA and should be available to start receiving manuscripts in early 2004 to prepare for issues published in 2005. Please note that the P&C Board encourages participation by members of underrepresented groups in the publication process and would particularly welcome such nominees. Self-nominations are also encouraged.

Search chairs have been appointed as follows:

- **Contemporary Psychology: APA Review of Books**: Susan H. McDaniel, PhD, and Mike Pressley, PhD
- **Developmental Psychology**: Joseph J. Campos, PhD
- **Psychological Review**: Mark I. Appelbaum, PhD

To nominate candidates, prepare a statement of one page or less in support of each candidate. Address all nominations to the appropriate search committee at the following address:

Karen Sellman, P&C Board Search Liaison
Room 2004
American Psychological Association
750 First Street, NE
Washington, DC 20002-4242

The first review of nominations will begin November 15, 2002. The deadline for accepting nominations is November 25, 2002.