COGNITIVE SCIENCE **20**, 457–497 (1996)

Shuttling Between Depictive Models and Abstract Rules: Induction and Fallback

DANIEL L. SCHWARTZ
Vanderbilt University

JOHN B. BLACK
Columbia University

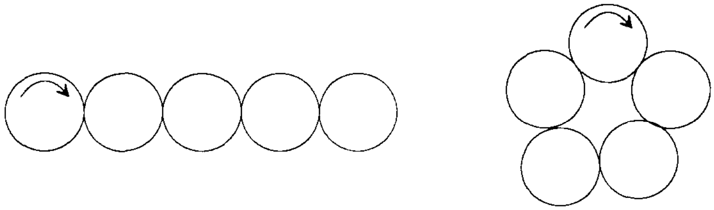
A productive way to think about imagistic mental models of physical systems is as though they were sources of quasi-empirical evidence. People depict or imagine events at those points in time when they would experiment with the world if possible. Moreover, just as they would do when observing the world, people induce patterns of behavior from the results depicted in their imaginations. These resulting patterns of behavior can then be cast into symbolic rules to simplify thinking about future problems and to reveal higher order relationships. Using simple gear problems, three experiments explored the occasions of use for, and the inductive transitions between, depictive models and number-based rules. The first two experiments used the convergent evidence of problem-solving latencies, hand motions, referential language and error data to document the initial use of a model, the induction of rules from the modeling results, and the fallback to a model when a rule fails. The third experiment explored the intermediate representations that facilitate the induction of rules from depictive models. The strengths and weaknesses of depictive modeling and more analytic systems of reasoning are delineated to motivate the reasons for these transitions.

When reasoning about physical systems, people sometimes experience the phenomenology of depicting the system's behavior in their imagination (Clement, 1994; diSessa, 1993; Hegarty, 1992). Because people do not experience this phenomenology throughout their reasoning, we assume that its onset reflects a functional shift in problem solving strategies. Beginning with this assumption, we consider why and when people use imagistic models to reason about a physical system, and how this imagery becomes related to

This research was supported by a Ben D. Wood Fellowship to the first author from Teachers College, Columbia University. Acknowledgments go to Micki Chi, David Krantz, Joyce Moore, Nancy Vye, and an anonymous reviewer. Parts of Experiment 1 were presented at the 1990 meeting of the American Educational Research Association, Boston.

Correspondence and requests for reprints should be sent to Daniel L. Schwartz, Department of Psychology and Human Development, Box 512 GPC, Vanderbilt University, Nashville, TN 37203. E-mail: <schwardl@ctrvax.vanderbilt.edu>

a. Open Chain Configuration **b. Closed Chain Configuration**



Five gears are arranged in a horizontal line. If you try to turn the gear on the left clockwise, what will the gear on the far right do?

Five gears are arranged in a circle so that each gear is touching two other gears. If you try to turn the gear on the top clockwise, what will the gear just to its right do?

Figure 1. A Visual Representation of Two Problems.

more abstract understanding. Our hypothesis is that one purpose of imagery is that it can serve as a surrogate for unavailable, or effortfully obtained, empirical evidence. This is not meant to imply that imagery evidence is necessarily veridical or capable of revealing previously unexperienced classes of phenomena. Rather, the hypothesis is that people often situate their reasoning in the structure of imagined experience when they cannot situate their reasoning in the structure of perceived experience. To examine the explanatory value of the hypothesis, we test two implications. First, imagery should be used to generate surrogate evidence at those points in time when people would perceive empirical evidence if they could. Second, much like reasoning over experience, people should be able to use the surrogate evidence of imagery to learn higher order rules. We test these implications by examining how people move between depictive imagery and rule-based reasoning depending on their current and evolving knowledge of gear interactions.

DEPICTION AS A SOURCE OF EVIDENCE FOR NOVEL PROBLEMS

If depictions of imaginary events serve as surrogates for unavailable evidence, then people should use depictions at those times when they would ideally use empirical evidence but for some reason cannot. For example, consider the gear problem shown in Figure 1a. As is characteristic of physical problems, there are several ways one might solve this problem depending on one's knowledge state and maturation (e.g., Bruner, 1966; Chi, Feltovich, & Glaser, 1981; Larkin, McDermott, Simon, & Simon, 1980; Metz, 1985; Weld, 1986; White & Frederiksen, 1990). One approach is to use a global level description or rule about the device (e.g., Hegarty, Just, & Morrison, 1988). For example, people could use a parity rule that states, "If there is an odd number of gears in a chain, then the first and last gears turn the same direction." Given sufficient confidence in the rule, there is little reason to resort to empirical evidence. Therefore, we should not expect people to imagine the system's behavior if they have a parity rule. A second approach is to rely on local rules about single interactions within the system. For example, "If two adjacent gears are touching, then they turn in opposite directions." Using this local rule, people could work through the gear chain, reasoning about pairs of gears and propagating the results from one local inference to the next (e.g., de Kleer & Brown, 1981). Again, it is not necessary to gather further evidence about gears, and therefore people should not imagine the gears in action. But what about the situation in which people do not have an appropriate global or local rule (cf. de Kleer & Brown, 1983)? For example, they may not have previously thought about pairings of gear motions. Or, they may have a parity rule that is too narrowly defined for the task at hand, as would be the case if one tried to apply a simple, open-chain parity rule to the locking problem in Figure 1b. In these situations, people would presumably gather empirical evidence if practical. Therefore, according to our hypothesis, it is in these situations, where people have inadequate rules and limited access to empirical evidence, that people should depict an imaginary model. They might, for example, simulate a left-hand gear turning clockwise against a right-hand gear to see what develops. Or, in the case of the closed gear chain, they might mentally model two adjacent gears each trying to turn clockwise.

The idea that people rely on mental models, imagistic or otherwise, to draw inferences about novel situations is not new. An original motivation behind mental model theory was to provide an explanation for how people reason about novel textual descriptions for which they do not have a pre-existing script or schema (Black & Bower, 1980; Collins, Brown, & Larkin, 1980; van Dijk & Kintsch, 1983). Rather than interpreting information into pre-existing knowledge structures, people use lower-level knowledge (e.g., p-prims, diSessa, 1988) to construct models that portray possible worlds consistent with the text (Bransford, Barclay, & Franks, 1972; Johnson-Laird, 1983; Mani & Johnson-Laird, 1982; McNamara, Miller, & Bransford, 1991). In the domain of physical and mechanical inference, the term mental model has referred to many different representational constructs, including rules. Cognitive scientists within the qualitative physics tradition, for example, have simulated model-based reasoning by employing predicates that can "represent continuous properties of the world by discrete systems of symbols." (p. 12, Forbus, 1990; but see Hendrix, 1973). In this article, however, we examine transitions between discrete symbolic rules and reasoning that mimics perceptual experience. Therefore, to maintain the distinction, we refer to *depictive models* when we mean models that attempt to reinstate perceptual experience.

Although not a central assumption for the current purposes, in our formulation (Schwartz & Black, 1996) a depictive model is much like analog imagery (Shepard & Cooper, 1982), except that it includes provisions for the representation of physical constraints (e.g., friction) as well as spatial ones (e.g., occlusion). Like analog imagery, a depictive model reveals an object's real-time, continuous changes by transforming a referent model according to internalized constraints (Shepard, 1994). Unlike analog imagery, depictive models can use represented physical constraints to coordinate the local interactions between imagined objects. For example, a friction constraint may help coordinate the imagined movements of two touching gear surfaces. Relevant to the current studies, the constraint-based formulation may help explain why people use depictions in situations of novelty. Deep knowledge embodied in a constraint may be difficult to formulate into a rule-based description of object behaviors, but it may reveal itself in the imagined object behaviors that it enforces. For example, Schwartz and T. Black (1996) studied inferences about nonrigid behaviors to demonstrate that a depiction can yield an inference that cannot be drawn otherwise. Participants saw two glasses that had different diameters, but had identical heights and levels of *imaginary* water. Their task was to determine whether the imaginary water in the two glasses would start pouring at the same angle of tilt, and if not, which glass would pour first. When participants simply looked at the glasses and made a qualitative judgment, they were rarely correct. However, when they closed their eyes and tilted each glass until the imagined water started to pour, they correctly tilted the narrower glass further. The success of the depictive strategy motivates why it may be adaptive to construct a depictive model as a surrogate for empirical evidence; namely, a mental depiction may provide access to sources of knowing unavailable to more discrete, rule-based reasoning. So, in the case of the gears, depictions may tap into constraint-based knowledge about touching objects (e.g., Funt, 1980; Parsons, 1994) that is not readily used in a more discrete, verbal formulation.

DEPICTION AS A SOURCE OF EVIDENCE FOR RULE INDUCTION

If people use depictions as surrogates for evidence, then people should be able to use their depictive results much as they do empirical results. In the current case, we examine whether people can use the patterns manifested in their depictions to induce higher order relationships. For example, imagine an individual who successively models five (e.g., Figure 1a), six, and seven gear problems. If these models serve as "quasimorphs" (Holland, Holyoak, Nisbett, & Thagard, 1986) of evidence, then the individual might induce a parity rule over the pattern: 5 = clockwise, 6 = counter-counter, 7 = clockwise. The idea that people can induce numerical rules from their mental depictions complements Johnson-Laird's observation that "by reflecting on the properties of relations represented in mental models, an individual may come to acquire a higher-order knowledge of them" (p. 191, Johnson-Laird, 1981).

The following experiments and their accompanying discussions document how the induction of abstract rules occurs over mental depictions. In addition to supporting the hypothesis that depictive models provide surrogate evidence, this documentation can begin to address the question of how nonnumerical, depictive representations can evolve into numerical, abstract representations. In the larger scheme of things, if one is to accept the plausibility of multiple forms of representation, it must be shown how and what information translates between these forms (e.g., Kosslyn, 1980). To simplify the task of identifying transitions between models and rules, we operationalize an abstract rule as a truth-valid assertion that relies on an articulated symbol system. By rules we do not mean production rules, although they may be the underlying form of an abstract rule. Rather, we mean a representation that depends on linguistic or numerical symbols. In the current case, a parity rule for the gears is abstract by virtue of its dependence on the mathematical relations of odd and even.

In addition to inducing an abstract rule that can characterize the patterns within their depictions, people should use this rule. To see why, consider a weakness of depictions and some strengths of rules. A demonstrated limitation of a dynamic, imagistic inference is that it can only simulate local interactions (Hegarty & Sims, 1994; cf. Kosslyn, 1980). Although depictive models of local interactions can be subcomponents of a larger model (Hegarty & Just, 1993), depictive models are not the same thing as the mental models that have been described for the understanding of complicated systems (e.g., Gentner & Stevens, 1983; Vosniadou & Brewer, 1993). To model a complex system through mental depiction, it would be necessary to model subassemblies of the system and chain the results of each simulation to adjacent subassemblies (Hegarty, 1992). When solving a 313 gear problem, for example, an individual may need to depict each pair of adjacent gears separately. This is a prohibitive task. In contrast, a parity rule allows for a quick derivation of the answer; 313 is an odd number, and therefore, the first and last gears turn the same direction. This example points out one strength of abstract rules. They tend to provide parsimonious methods for solving classes of well-defined problems.

A second motivation for shifting to a rule-based representation is that people can make inferences based on the structure of the rule's symbolic domain. For example, once formulated in a logical fashion, a rule can be used to reason about relations like contradiction. Imagine applying an open-chain parity rule to a circular, five gear problem (Figure 1b). The parity rule makes it possible to use contradiction to think about the problem. For example, the fifth (last) gear is odd and should turn clockwise, but this gear is also the second gear counting the other direction, and therefore should turn counter-clockwise. These contradictory predictions can indicate that something is amiss with the original parity rule, although they cannot determine the correct outcome. Solely on the basis of this contradiction and without further physical or imagined evidence, people could just as well infer that the gears would jump off the table as they could infer that they would jam. Nonetheless, the example does show how the structure of a symbolic domain can support forms of reasoning not available to mental depiction. One's imagination, like an experiment, reveals determinate "factual" results, not contradictions. Only by virtue of one's theoretical or rule-based interpretation does a fact have something to contradict.

SHUTTLING BETWEEN DEPICTIVE MODELS AND ABSTRACT RULES

To show that people use depictions as surrogates for evidence, the following experiments document how people shuttle between depictive models and abstract rules. All three experiments used verbally presented gear problems. Our thought was that the gear problems would yield a crisp example of the movement between model and rule that, in other tasks, may occur through less overt, and more frequent, shuttling within a single problem. In each experiment, people solved a series of similar problems (e.g., open-chain problems) and then confronted new, but related problems (e.g., closed-chain problems). Given that participants never actually see any gears, the surrogate-for-evidence hypothesis leads to three general predictions. The first prediction is that people tend to use depictive models when a situation is novel with respect to their current body of rules, as would be the case when participants are first introduced to the gear problems. The second prediction is that people can induce a rule from a pattern of depictive results, as should occur if participants solve several open-chain problems in a row. The final general prediction is that people fall back to depictive modeling when their rules become inadequate, as should be the case when participants induce an open-chain rule and are then given a locking, closed-chain problem.

The series of gear problems created an unusual induction task. In addition to creating a situation where a correctly induced rule would subsequently fail, the task differed from previous induction research in three ways. First, there was no physical stimulus over which people could induce a rule. This differentiates the current work from examinations of how people coordinate experiential and theoretical knowledge during induction (e.g., Klahr & Dunbar, 1988). Although we could have allowed participants to use real gears for this task, people can solve the problems without the gears. Reasoning and learning in the absence of physical evidence is not atypical (e.g., using a text). People manage to learn from their own thoughts and we are examining this process.

Second, the induction of a parity rule can be distinct from the accurate solution to the problems. In concept attainment and insight research, the induction or insight is measured by whether and when a participant solves the given problem (e.g., Kaplan & Simon, 1990). An example of this paradigm is a task for which the participant must induce the rule that is generating a sequence of playing cards (e.g., odd hearts, Laughlin & Shippy, 1983). In this paradigm, problem solution is predicated upon induction. For the gear task, the induction is predicated upon problem solution. People can infer an accurate solution without ever discovering the parity rule. Thus, unlike concept attainment and insight tasks, the inductive insight that leads to the parity rule is not required to solve the problems nor is the participant necessarily aware that such a rule exists. This situation parallels educational settings in which we want students not just to solve problems, but also to induce rules that can apply across a number of problems.

Third, the induction task differs from prior studies that used fairly abstract tasks (e.g., Tower of Hanoi, Simon & Lea, 1974). Abstract tasks are not naturally suited to the question of how perceptual representations of physical causality become related to more formal, mathematical representations. Using the problem space representation typical of tasks like the Tower of Hanoi, Metz (1985) analyzed representational transitions in reasoning about gears. Although a thorough investigation, the emphasis on a problem space analysis downplayed the shifts between perceptual and analytic strategies that are the focus here.

EXPERIMENT 1

In Experiment 1, participants heard 12 gear problems. The first six problems described gears in an open chain. According to our hypothesis, participants should depict the first few problems to generate evidence about interacting gears, assuming they do not already have an applicable rule. Subsequently, they should use this evidence to help induce a parity rule. Prior induction research has shown that people often require two examples from which to abstract structural patterns (e.g., Gick & Holyoak, 1983). Accordingly, rule induction should occur around the third problem and rule use should occur for the following three problems. The second six problems described gears in a closed chain. If participants induce a parity rule from the open-chain problems, it should fail when applied to a locking problem. According to our hypothesis, participants should model these problems because they cannot rectify their parity rules on the basis of mathematics or logic alone. They should require new evidence about adjacent gears turning in the same direction. Once participants solve three closed-chain problems, they should induce another parity rule that takes into account the locking behavior of an odd, closed chain.

In this experiment, latency and gestural data were used to indicate model and parity rule use. Rule-based solutions should be relatively fast. For these problems, a parity rule simply requires noting the direction of the initial gear and determining the oddness or evenness of the gear chain. In contrast, modeling is a sluggish way to solve problems that involve several components and motions. Although depictive models can operate very quickly (e.g., Schwartz, 1995b), they can only model a limited number of events simultaneously (Hegarty & Sims, 1994). So, for example, a participant might model the first and second gears, then the second and third gears, then the third and fourth gears, and so forth. As a result, modeling latencies should be longer than rule-based latencies for this task. For example, compare the time it takes to model 20 gear motions with the time it takes to determine that 20 is an even number.

The second data source was hand gestures. Pilot work revealed that these problems occasion the frequent use of hand gestures. Several pilot participants, for example, splayed their fingers on both hands and then rotated them inward mimicking the behavior of meshing gears. The possibility that people use these gestural models to solve the current problems brings the importance of depictive models into theoretical relief. The models people create with their hands straddle the boundary between physical and mental models. From one perspective, the hand movements are a method of extending one's thinking into the environment for further reflection (Resnick, 1987). From another perspective, the hands create a representation that brings the physical environment into one's thinking (Scribner, 1984). Hand movements, by being physically instantiated mental models, highlight the idea that people may reason similarly over internally and externally generated phenomena (Schwartz & T. Black, 1996).

The hand gestures in focus here are not the same as the speech accompaniment gestures often investigated in social psychology (e.g., Ekman & Friesen, 1972). Whereas speech accompaniment gestures, like fist pounding and rhythmic beats, tend to have metaphorical and pragmatic relations to discourse semantics (but see Alibali & Goldin-Meadow, 1993), the current gestures represent their referents deliberately and directly, as in molding one's hand into a gear shape (McNeil, 1987). Unlike speech accompaniment gestures, these referential gestures can stand independently of verbal processes (Saltz & Donnewerth-Nolan, 1981). For example, many participants in the following studies silently stared at their moving hands. Hand gestures that take the form of their referents have been shown to influence semantic sensibility judgments (Klatzky, Pellegrino, McCloskey, & Doherty, 1989). Given that referential gestures can play a role in semantic tasks without verbal mediation, it is reasonable to suppose that these representations can also support inferences. It is important to note, however, that this article does not offer proof that hand gestures play a functional role in solving the gear problems. Although there is evidence that bodily action facilitates imagined transformations in navigational tasks (Rieser, Garing, & Young, 1994) and object rotation tasks (Rieser & Schwartz, in preparation), one cannot be sure that gestures help people solve the causal problems employed here. Perhaps the gestures are simply correlated accompaniments to a "deeper layer" of representation. However, even if this latter case is true, gestures can open a fresh window on model-based reasoning by virtue of their correlation with deeper layers of representation.

Gesturing behavior is a potentially useful source of evidence that people are reasoning depictively. One strength of the data is that many people gesture spontaneously. This makes for an unobtrusive source of data. Another strength is that hand gestures may directly reflect spatial reasoning. As such, gestural measures may be better suited to capturing the spatial simulations of a depiction than verbal measures. Moreover, the data source directly indicates modeling. Unlike analog imagery, where the depiction occurs internally, one does not need to infer the existence of a depictive model when using gestural data; the referent model is in plain view.

Method

Participants

Twenty-four paid graduate students from Columbia University participated.

Design

Twelve gear problems (three short open-chain, three long open-chain, three short closed-chain, and three long closed-chain) created a 2×2 within-subject design. The *configuration* factor represented the open and closed chains. Participants always heard the open-chain problems first. There were six problems in each configuration. The *block* factor split each set of six problems into three problems per block. The first blocks in each configuration used randomly selected problems from three to six gears in size. The second blocks used three problems randomly selected from the range of five to nine gears. No chain length was repeated within a configuration. First block chains were shorter than second block chains to ensure that long latencies for the first blocks were not due to longer chain lengths.

Procedure

The participant sat facing a video camera and was screened from the experimenter. The experimenter read, "All of the following problems involve reasoning about gears. You should assume that each gear is touching its closest neighbors. I am only interested in your answer and do not want you to explain what you are thinking. If you have a question about the problem I will only re-read it. Be sure of your answer. If you get it wrong, I will tell you so and you will have a second opportunity to solve the problem. After that I will tell you whether you are right or wrong and go onto the next problem." If participants exhibited confusion over the type of gears (e.g., a car shift), the gears were compared to quarters, side-by-side on a table. There were no visual stimuli. The experimenter then read, "[Five] gears are arranged in a horizontal line, if you try to turn the gear on the far left clockwise, what will the gear on the far right do?" Participants worked without prompting until they reported an answer. If the answer was correct, the experimenter read the next problem. If the answer was incorrect the participants were given a second try, because we did not assume that participants would necessarily reason that if the gear did not turn one way, it must turn the other. A participant might try to remodel the problem instead. After participants completed the open-chain problems, the experimenter stated, "In the following problems, you are to assume that each gear is touching two other gears." The experimenter then said, "[Five] gears are arranged in a circle, if you try to turn the gear on the top clockwise, what will the gear just to its left do?" As debriefing, participants either taught a naive individual how to reach correct answers, or explained the reasoning behind their answers.

Coding

A primary coder measured how much time each individual spent making rotation gestures for each problem. To count as a rotation gesture, a motion had to have at least a 90° circular movement of an arm, hand, or finger. The coded duration of the gesture only included active movements. This avoided over-estimating the amount of dynamic modeling because it excluded periods when individuals left their hands suspended in mid-air during reflection or speech. An independent judge, blind to the hypotheses, coded gesturing times for each problem for four randomly selected participants. The independent judge and the primary coder had strong agreement; $R = .99$. A second type of coding was whether the individuals spontaneously mentioned a parity rule during problem solving or debriefing. Although parity rules may vary in generality, they all involve a parameter for the oddness or evenness of the number of gears. Accordingly, a mention of "odd" or "even" in conjunction with a statement of gear direction is reasonably indicative of a parity rule at play. There were no coder disagreements. Finally, problem-solving latency was measured as the time between the last word of the problem presentation and the first word beginning a correct or second answer.

Results

Twelve of the 24 participants performed above chance on their first answers to the problems in both configurations. Of the 12 individuals who performed

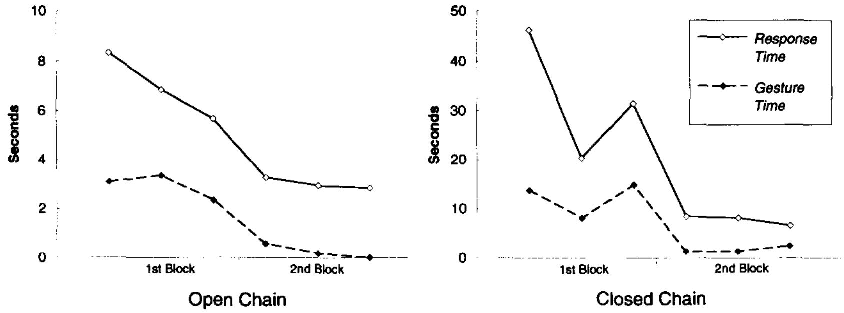


Figure 2. Response and Gesturing Times by Problem Order (Experiment 1). Please note that the open- and closed-chain graphs use different scales. Refer to the text for an explanation of the similar patterns but differing magnitudes.

below chance, four could not solve either the open- or closed-chain problems, and eight were unable to solve the closed-chain problems.[1] Because of the focus on transitions between model and rule, we confine our analysis to the 12 successful problem solvers. Figure 2 displays the problem-solving latencies and the time spent making rotational gestures. Both measures dropped to relatively stable levels after the third problem in the open chain, jumped with the introduction of the closed-chain problems, and then dropped again to stable levels after the third closed-chain problem. This suggests that participants initially modeled the problems for each configuration until they induced a more efficient solution rule.

To analyze the data, we compared each individual's average per problem latency and gesturing time to reach a final answer (i.e., a correct answer or an incorrect second answer) for each block of three problems. The top two rows of Table 1 show the averages and variability across individuals. Each individual's four latency and four gesturing means were used to test the within-subject factors of configuration and block. The reported univariate significance values are adjusted according to the Huyhn-Feldt (HF) correction value if necessary (Cliff, 1987). There was a main effect of the problem

[1] The four participants who could not solve the open-chain problems constructed unanticipated models from the verbal descripiton of the gear problems. For example, during debriefing two of these participants mentioned that they thought of the gears as though they were tires on a truck. A third participant said she thought of the gears as stacked on top of one another. They did not construct models in which the gear surfaces interacted with one another. The other eight participants, who could solve the open-chain problems but not the closed-chain problems, had gesturing and latency patterns similar to the successful problem solvers for the open chains. However, they were never able to infer that the gears would lock and consequently never arrived at an adequate parity rule for the closed-chain problems. These subcriterion performances show one prerequisite of a successful depiction; namely, one must first determine what situation to depict. The participants who could not solve the open-chain problems evidently constructed models for a different problem. The participants who could not solve the closed-chain problems never tried to model two gears that were turning the same direction.

TABLE 1
Average Per Problem Times for Each Problem Block (Experiment 1)

	Open Chain		Closed Chain	
	1st Block	2nd Block	1st Block	2nd Block
Response Time (s)	6.9 (3.9)[a]	3.0 (2.9)	32.6 (24.1)	7.7 (5.8)
Gesturing Time (s)	2.9 (3.9)	0.3 (0.6)	12.2 (17.2)	1.7 (2.5)
Gesture Density (GT/RT)	.36 (.37)	.04 (.08)	.34 (.30)	.19 (.19)

[a] Standard deviation within parentheses across participants.

block for both measures; $F(2,10)=8.35$, $p<.01$, $HF=.66$. For both configurations, the latencies and gestures decreased on the second block of three problems; $F(1,11)=16.81$, $p<.01$; $F(1,11)=5.47$, $p<.05$, respectively. The main effect of configuration was also significant; $F(2,10)=7.48$, $p<.05$, $HF=.69$, with closed-chain problems leading to longer overall latencies; $F(1,11)=16.45$, $p<.01$, and more gesturing; $F(1,11)=6.3$, $p<.05$. The block by configuration interaction was also significant; $F(2,10)=4.2$, $p<.05$, $HF=.70$. This was primarily due to the latencies which exhibited a larger drop from the first to second block of problems for the closed-chain than for the open-chain; $F(1,11)=9.2$, $p<.05$. The block by configuration interaction was marginal for the gesturing times; $F(1,11)=3.95$, $p<.1$.

Not only did people gesture longer during the first blocks of each configuration, but as shown in Figure 3, they also spent a greater proportion of their problem-solving time making gestures. We call this proportion the *gesture density*. The last row of Table 1 shows the average gesture density per problem for each block of three problems. There was a main effect of block showing that there was a greater gesture density in the first block compared to the second in each configuration; $F(1,11)=9.44$, $p<.05$. There was no main effect of configuration; $F(1,11)=1.17$, but there was a block by configuration interaction; $F(1,11)=5.05$, $p<.05$. The interaction reflects the greater gesture density in the second block of the closed chain compared to the second block of the open chain. One interpretation of this interaction is that several individuals had not induced a parity rule by the end of the third problem for the closed chain, and therefore continued to model the problems until they induced a rule for the closed chain. For the first block of the open chain, 14% of the problems were answered incorrectly on the first try. However, for the first block of the closed chain, 39% of the first answers were incorrect. This means that several participants had only recently learned of the correct answer by the time they reached the second block of the closed-chain problems. Thus, they may not have had time to induce the parity rule over a set of stable answers. This interpretation may also explain why the problem-solving latencies for the second block of the closed chain were longer than the second block of the open chain. Some

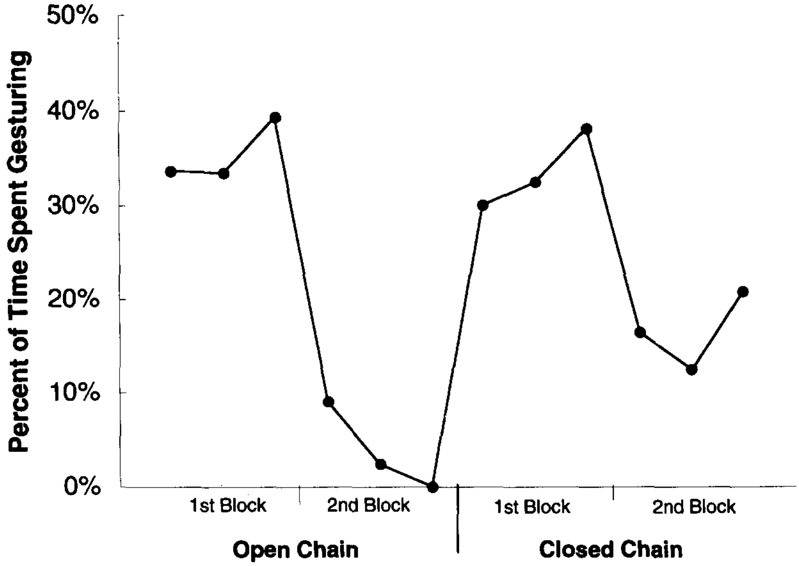


Figure 3. Gesture Density by Problem Order (Experiment 1). The proportion of problem-solving time using rotating hand movements.

individuals were still modeling the problems to ensure the correctness of their answers. By looking at the final problem of the closed-chain configuration, instead of averaging over the block of three problems, we may see that problem-solving latencies for the closed chain eventually showed the brevity expected of parity rule use. For the final problem, if we exclude the 40 s latency of one person, the average response time was 3.5 s (4.1 *SD*), very much in accord with the latencies for the post-induction, open-chain problems.

As supporting evidence that participants had induced a parity rule, 10 of the 12 individuals who performed above chance spontaneously mentioned a version of the parity rule during debriefing. An 11th person stated the number of gears, "same" or "opposite," and then "clockwise" or "counterclockwise." The fact that this person stated "same" or "opposite" before mentioning a direction of motion, suggests that he resolved the problem on the basis of the odd/even pattern. The remaining individual who did not mention the parity rule did not appear to have induced a parity rule. As the chain lengths increased, his responses took longer and the time spent modeling increased as he iterated, "clock. . .counter. . .," down the longer chains of gears.

Discussion

Our interpretation of the results is that the rise and fall in dependent measures reflected a shuttling between depictive models and abstract parity rules. When participants first encountered the open chain, they had no rule in place and depicted the gears to solve the problems. As a result, the problem-solving and gesturing times were relatively long. After having successfully modeled three problems, individuals had sufficient data to induce a parity rule. Consequently, latencies became shorter and gesturing activity disappeared by the sixth problem. Supporting this interpretation, 10 of the 12 successful participants reported a parity rule. Their initial rules, however, were insufficient for solving the locking, closed-chain problems. To discover the behavior of gears in a closed chain, participants fell back to a depictive strategy as reflected in the increased latencies and gesturing. Response times were especially long because people had difficulty discovering that the gears would lock; enough so that eight individuals who had correctly solved the open-chain problems were unable to solve the closed-chain problems. Those participants who did successfully solve the closed-chain problems were able to induce an improved parity rule, as indicated by the steep drop in latencies and gesturing that decreased for the final closed-chain problems. Latencies and gesturing for the second block of closed-chain problems were greater than the comparable open-chain problems, perhaps due to the increased difficulty of discovering the locking outcome. Because several participants did not determine the locking behavior until the later problems, they did not have an opportunity to induce a pattern of behavior over the first three closed-chain problems.

There are four primary alternatives to our interpretation. The first two question the evidence of rule induction and rule use. The second two question whether the modeling for the first closed-chain problems was a generalizable example of fallback modeling. We first consider whether participants induced a rule. Given the current evidence, one might argue that participants became more proficient modelers over the trials and never used a rule-based representation. For example, they may have become practiced enough with their models that they could sustain them solely through internal imagery, thereby explaining why gestures diminished. The complete reliance on internal imagery might also explain why latencies diminished, if one presumes that their internal imagery would be quicker than deploying their hands. One problem with this practice-effect interpretation is that the latencies and gestures did not diminish smoothly. Figure 3 shows that the gesture density shifted dramatically between the third and fourth problems in both configurations. Nonetheless, this precipitous drop only demonstrates that people stopped using gestural models rather suddenly, not that they started using rules.

The second alternative interpretation is that participants had parity rules all along. For the first few problems in each configuration, participants may have employed both gestural models and parity rules. It was not until their rules had received sufficient confirmation over a few trials that they were willing to rely on them exclusively. Under this interpretaion, rule use should not be viewed as a result of induction, but rather as a result of gaining ascendancy in a strategy competition. Alternatively, participants may have applied rules all along, it was just that their initial rules were faulty and could not be relied upon exclusively. In this case, participants did not induce parity rules, but rather modified or pruned faulty rules using feedback or their own modeling.

To counter these two alternative interpretations it is necessary to show that people actually induce a rule. Two lines of verbal evidence, not collected in the current study, could be relevant. One type of evidence would document the point of induction directly, perhaps through think aloud protocols, and would show how the measures shift about this point. The second type of verbal evidence could show that rule-based behaviors primarily appear after induction. If participants were encouraged to verbalize their thoughts, there might be a shift in language that complements a shift in representations. When participants model the problem they might make references primarily to their models as in, "This one goes that way." However, once participants induce a parity rule, they might make references primarily to the quantities in the problems as in, "Three gears is an odd number." Experiment 2 developed the verbal evidence pertinent to the induction of a rule and the shift to rule-based reasoning.

The third alternative to our interpretation accepts the idea that participants induced a rule from their models, but disagrees with the fallback interpretation of the closed-chain modeling. By this and our interpretation, participants used models when their rules were insufficient for the task. However, for the purpose of clarifying how people use model evidence to inform their rules, we wish to make a further differentiation of the ways in which one's body of rules may be inadequate for a given task, such as situations where one has no rules (i.e., the first open-chain problems) and situations where one's rules fail (i.e., the first closed-chain problems). The alternative interpretation does not embrace the differentiation of the open- and closed-chain modeling. Instead of falling back, people might have modeled the closed-chain problems for the same reason that they modeled the open-chain problems; namely they were both novel problems for which the participants had no rule.

To support the fallback interpretation, one might demonstrate that people try to apply their rules, and then after these rules fail, people begin modeling the problem. For example, several participants in the preceding experiment clearly noted the contradictory, clockwise and counter-clockwise, predictions of their open-chain rule when solving an odd, closed-chain problem. However, definitive evidence like this may be hard to develop because people could silently consider and reject their rules without external manifestation. An alternative approach is to show that people include knowledge of the open chain in their closed-chain rules. Although this does not directly prove fallback modeling, it can show that the closed-chain problems were not novel in the same sense as the open-chain problems.

The final alternative interpretation questions whether the modeling found for the first locking problems demonstrates the generality of fallback modeling. According to this alternative, the modeling for the closed-chain problems was due to the excessive difficulty of the locking problems, especially after people had developed a nonlocking set. In this case, the claim is that the fallback modeling demonstrated for the locking problems is a special case and does not generalize to less problematic situations. To evaluate this alternative, one can examine whether people model for a new problem that is not as difficult as the closed-chain problem. For example, after people induce a parity rule for the open-chain, they could be asked to solve an open-chain problem in which the initial gear turns counter-clockwise instead of clockwise. If they model for this simple variant, then the unique difficulties of a locking problem are not requisite for fallback modeling.

EXPERIMENT 2

The current experiment was designed to corroborate Experiment 1 and to gather verbal measures of rule induction and rule use. It was also designed to examine how different types of rule insufficiencies influence the shuttling between models and rules. There were several predictions at test. As before, we predicted that people model in situations where they do not have adequate rules, and that people can apply rules induced from their model simulations. To complement the latency and gestural data, verbal measures were used to evaluate rule induction and subsequent use. A new prediction in this experiment was that in situations of rule failure, as compared to situations of novelty, people induce new rules that include elements of their original rules. We also tested the hypothesis that people generally fall back to modeling in situations where their rules are too narrow, not just in the unusual case of the locking problems.

To generate the necessary data, there were two primary modifications to the first experiment. To encourage verbal production, participants worked in dyads. We employed pairs of participants, instead of think aloud directives, because pilot work had shown that individuals tend to fall silent as they work through their model simulations. Moreover, a member of a dyad would presumably be quick to share a discovered parity rule, thereby indicating the approximate point of induction.

The second modification was to append *change* problems to the original six problems in each configuration. The top of Figure 4 shows the original problems used in Experiment 1 and here. The gear that initiated the movement always turned clockwise and was on the left of the open chain or on the top of the closed chain. The bottom of the figure shows the new change problems added to each configuration. For one of the problems, the initial gear turned counter-clockwise. For the other problem, the initial gear was

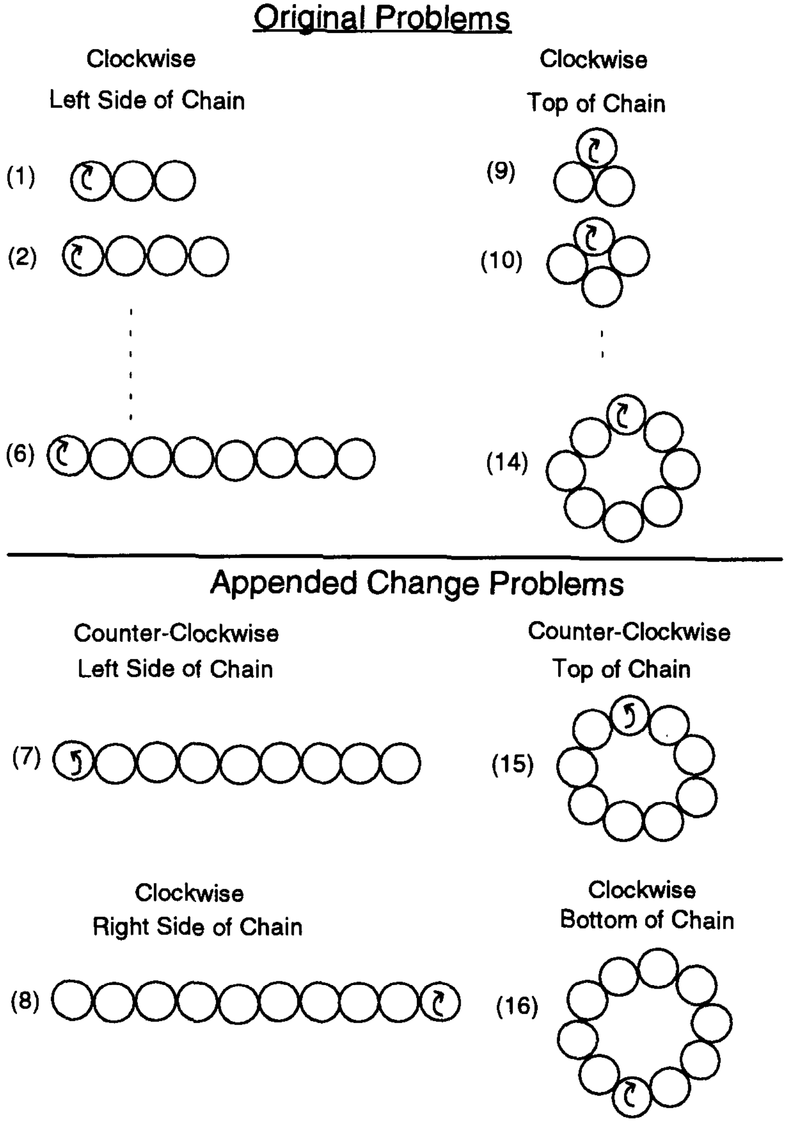


Figure 4. Appended Change Problems for Experiment 2. Two problems that changed the direction and position of the "first" gear were added to each configuration. Parenthetical numbers indicate the presentation sequence.

positioned on the right or bottom of the gear chain. These changes should make it so the participants' recently induced parity rules would not be directly applicable to the problem, and thereby reveal whether people fell back to modeling for problems that were not as problematic as the locking problems.

Moreover, if participants modeled the open-chain change problems, one can determine whether these results were carried over to their closed-chain parity rules. If participants simply applied a closed-chain parity rule to the change problems without modeling, then one may assume that the closed-chain problems were not completely novel because participants incorporated information from the open-chain problems.

Method

Participants

Twenty-four paid graduate students from Columbia University were randomly paired.

Design

The open- and closed-chain configurations each had eight problems that were presented in ascending order from three to ten gears. Figure 4 provides a schematic. We separate the experimental design that replicated Experiment 1 from the design that used the *change* problems to examine fallback modeling. Replicating Experiment 1, a block by configuration design compared behaviors during the first six problems in each configuration. Unlike Experiment 1, the point of induction was measured by the statement of a parity rule. This provided an empirical, within-subject separation of pre- and post-induction problem blocks. The second experimental design focused on the effects of the change problems. The design for this part of the experiment was configuration by change/no-change problem. Behavior on the sixth problem in each configuration (i.e., the last no-change problems) was used as a measure of stable rule-based behaviors. It was compared to behavior for the seventh and eighth problems within each configuration which changed the rotation direction and position of the initial gear, respectively.

Procedure

Dyad members, screened from the experimenter, faced each other over a low table and had several minutes to become acquainted. They were told that they would hear a number of problems, that they had to agree before reporting an answer, and that they had to remain in their seats. They then heard the same instructions as Experiment 1 with the modification that they continue solving each problem until they gave the right answer.

Coding

The appropriate unit of analysis was the dyad, not the individual. Because of shared reasoning and representations (e.g., one partner models and the other tracks the number of gears), separating the data according to individuals would be a difficult, if not impossible, task (cf. Schwartz, 1995a).

For the current experiment we simplified the gesture coding process by tallying the frequency of gestural initiations rather than timing their durations. A continuous movement of several rotations counted as one rotation initiation. This simpler coding method was sufficient for testing the predictions.

A primary coder tallied three types of language: exophoric references, quantitative words, and parity rule statements. Exophoric references like 'this' or 'that' should provide a convenient verbal measure of external modeling, because unlike anaphoric references, they point to referents that are external to the surrounding sentential context (Halliday & Hasan, 1976). The expression, "*This* goes *that* way," contains two exophoric references, but the expression, "Odd gears. . .*this* is like the earlier one," contains no exophoric references because the "this" makes an anaphoric reference to the odd gears in the preceding clause. Quantitative words referred to the cardinal properties of the gear chain. For example, the expression "*Nine* gears; that's an *odd* number," has two quantitative expressions. This category of expression also included questions like, "How many?" Quantitative expressions should be good indicators of rule use as participants became primarily concerned with the numerosity and parity of the problem. Ordinal uses of number words were not coded as quantitative expressions. For example, the enumerative sequence, "Gear one goes this way, gear two goes like this, and the third one goes like this," does not contain any quantitative references according to the coding scheme. Ordinal expressions were excluded from the quantitative category, because they reflected people's reasoning down a series of gears.

An independent judge coded three randomly selected dyads. The correlation between the primary coder and the independent judge was $R = .95$ for the gestures. Coding of parity rule statements, exophoric words, and quantitative words had no disagreements. The primary coder and judge had 94% agreement on judgments of gesture and reference simultaneity.

Results

Replicating Experiment 1—Gesture and Latency Data

Of the 12 dyads, nine mentioned a parity rule for both configurations, one dyad was unable to solve the closed-chain problems, and two induced an alternative procedure for solving the problems (see below) and never mentioned a parity rule. Because our predictions focus on transitions between models and rules, we analyze the nine dyads who induced the parity rule. Figure 5 shows that four of the nine dyads made errors and that their errors were confined to the change problems and the first locking problem. For the open-chain problems, the fourth problem was the mode of when the dyads first mentioned the parity rule with the average number of completed problems at 3.2 (1.4 *SD*). For the closed-chain problems, eight of the dyads

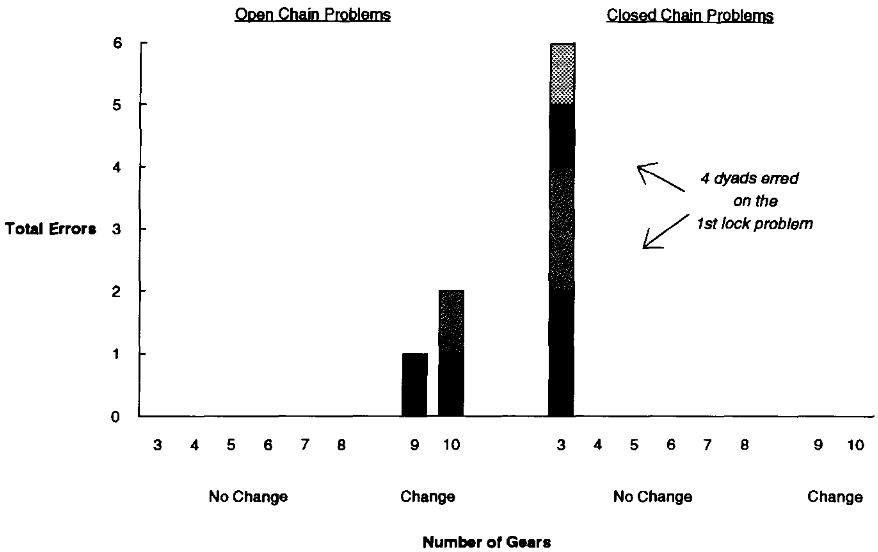


Figure 5. Errors for Dyads who Induced a Parity Rule (Experiment 2). The four different shadings for the graph represent each of the four rule-inducing dyads who made errors and the problem(s) on which they made an error.

explicitly stated a parity rule during the problem solving. The ninth pair communicated more tacitly and did not explicitly mention a closed-chain parity rule until debriefing. Problem-solving latencies may be used to infer when this ninth pair induced the rule. The closed-chain latencies for the eight, rule-mentioning dyads dropped six fold from a mean of 77.3 s (84.3 *SD*) on the first rule-mentioned problem to a mean of 12.3 s (27.0 *SD*) on the immediately following problem; $F(1,7) = 7.36$, $p < .05$. This result suggests that the dyad who did not explicitly mention the parity rule had actually induced the rule between the fifth problem (192 s) and the sixth problem (17 s). This eleven-fold drop in latencies did not elevate for subsequent problems. Including all nine dyads, the third problem of the closed-chain was the mode of rule induction and the average number of completed problems was 2.7 (0.7 *SD*).

Figure 6 shows that the latencies and gestures dropped after dyads explicitly mentioned a parity rule. At this point we compare measures from the first six problems in each configuration; the change problems are treated separately below. Because response times were used to infer the point of induction for one dyad, we only consider gestures statistically. The analysis contrasted the average number of rotational hand movements per problem, before and after dyads stated a parity rule for each configuration. This yielded four within-dyad measures capturing the factors of pre-/post-induction and configuration. There was a reliable drop in gestures after the state-

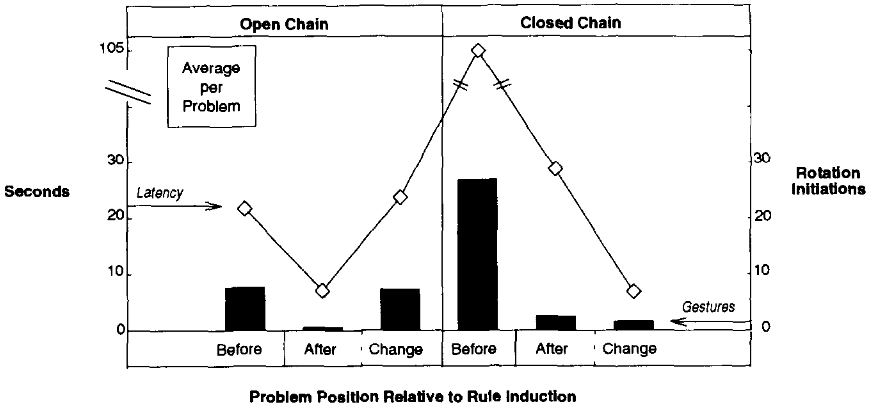


Figure 6. Response Times and Gesture Frequency (Experiment 2).

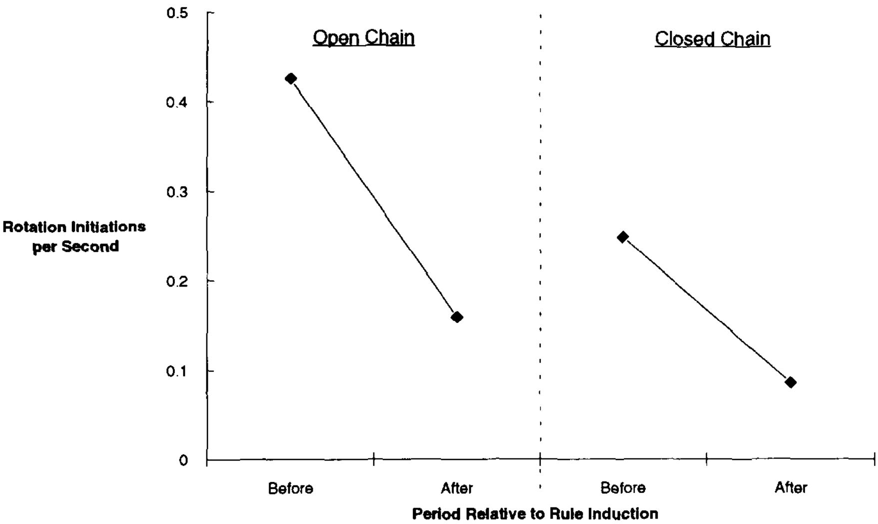


Figure 7. Gesture Density for Non-Change Problems (Experiment 2).

ment of a rule; $F(1,8) = 16.8$, $MSe = 131.59$, $p < .01$. There was also a main effect of the problem configuration with more gestures occurring overall in the closed-chain problems; $F(1,8) = 11.6$, $MSe = 86.8$, $p < .01$. As may be seen in Figure 6 and was supported by the pre-/post-induction by configuration interaction; $F(1,8) = 8.4$, $MSe = 82.5$, $p < .05$, this effect was primarily due to the larger number of gestures prior to rule induction in the closed configuration. The effects of induction and configuration were replicated analyzing the gesture densities shown in Figure 7. There was a pre-/post-induction effect; $F(1,8) = 46.5$, $MSe = .01$, $p < .01$, a configuration effect;

TABLE 2
References by Gestures (Experiment 2)

Simultaneous Gestures	Exophoric Words	Quantitative Words
Rotating	297 (82%)	79 (25%)
Not Rotating	65 (18%)	232 (75%)

$F(1,8) = 8.9$, $MSe = .02$, $p < .05$, and a pre-/post-induction by configuration interaction; $F(1,8) = 5.86$, $MSe = .00$, $p < .05$.

The pre-/post-induction by configuration interactions merit further explanation. Figure 7 shows that the gesture densities for the pre-induction, closed-chain problems were relatively low compared to the pre-induction, open-chain problems. Yet, Figure 6 shows that the frequency of gesturing was greater for the pre-induction, closed-chain problems compared to the pre-induction, open-chain problems. The videotapes reveal that these opposing interactions were an artifact of the extremely long latencies for the first closed-chain problems. During the first locking problems, perhaps due to frustration, dyads typically spent substantial time off-task by questioning the purpose of the experiment and discussing incidental matters. Although we do not quantify the off-task behaviors for the different problem blocks, our interpretation of the low gesture density for the pre-induction, closed-chain block is that it does not accurately reflect the proportion of actual problem-solving time spent using gestures.

A second aspect of the data that merits discussion is the relatively high latency and gesture frequency that followed rule induction in the closed-chain problems. This occurred because the member of the dyad who discovered the rule often explained the rule and locking behavior to the other member. For example, one individual said, "When there are an odd number of gears, these two top gears are both trying to turn this way. But touching gears can't both turn this way. See how they jam each other." Once both members of the dyad understood the rule and the locking behavior, the latencies and gestures dropped to minimal levels, paralleling the case for the open-chain problems. For example, on the sixth problem of each configuration, the average latencies were less than 5 s and gesturing times were less than 1 s.

Verbal Evidence of Rule Induction and Rule Use

Table 2 shows that exophoric references (e.g., "*this* one") were typically accompanied by rotational gestures, whereas the quantitative references were not; $\chi^2(1) = 217.77$, $p < .01$. This indicates that exophoric references were made with respect to dynamic gestural models. Figure 8 shows the rate of language use per second. For both configurations, dyads used exophoric references prior to rule induction and quantitative references after

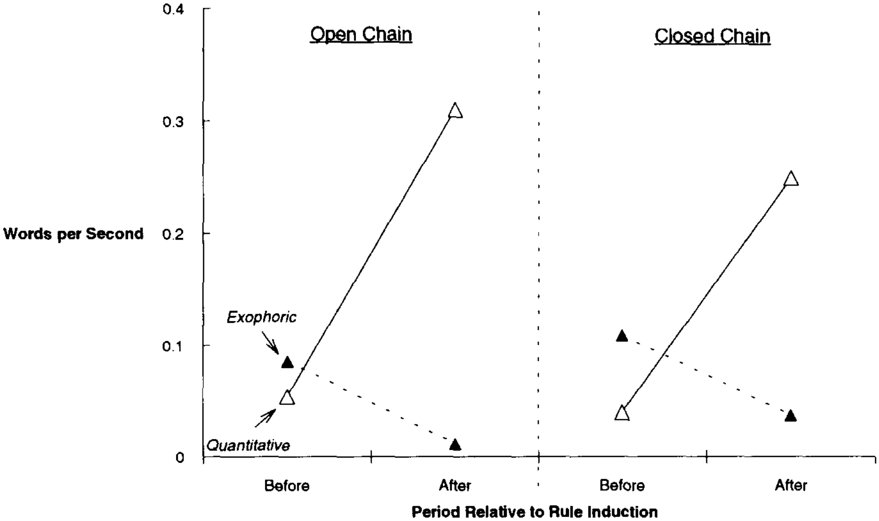


Figure 8. References per Second (Experiment 2). The average number of exophoric and quantitative references per second split by the statement of the parity rule (nonchange problems only).

induction. To show this cross-over statistically, we computed each dyad's rate of exophoric and quantitative word use per second before and after induction in both configurations. The eight measures created a reference type by pre-/post-induction by configuration within-subject design. There was a reference type by induction interaction; $F(1,8) = 60.9$, $MSe = .01$, $p < .01$. The interaction was replicated using number of expressions per problem; $F = 15.8$, and total number of expressions; $F = 16.1$.

Effects of Change Problems on Rule and Model Use—Gesture and Latency Data

Figure 9 shows the effect of the two change problems on latency and gesturing. For the open-chain change problems both measures of modeling increased compared to the immediately preceding problem. In contrast, there was little evidence of a fallback to modeling for the closed-chain change problems. A change/no-change by configuration within-subject design compared dyad latencies and gestures for the last prechange problem to the average of the two change problems in each configuration. The key two-way interaction was significant; $F(2,7) = 5.66$, $p < .05$, $HF = .65$. Compared to the other three cells in the design, the open-chain change problems led to elevated gestures and latencies; $F(1,8) = 7.7$, $MSe = 11.9$, $p < .05$; $F(1,8) = 11.2$, $MSe = 82.4$, $p < .05$, respectively. The fact that dyads fell back to modeling for the open-chain change problems supports the claim

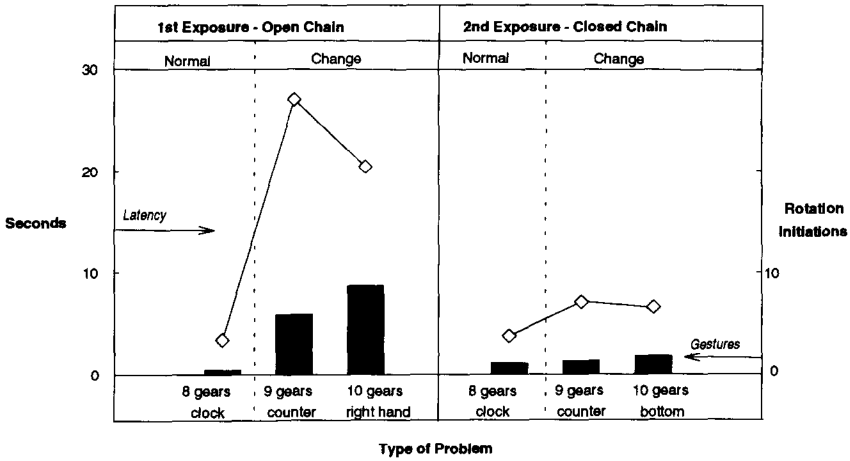


Figure 9. Gestures and Latencies for the Change Problems (Experiment 2).

that people generally model in situations where their rules are too narrow, and not just in situations that introduce unsuspected outcomes (i.e., locking). The fact that the dyads did not model for the closed-chain change problems supports the claim that participants did not treat the closed-chain problems as completely novel. The relevant information for solving the closed-chain change problems must have come from the open-chain problems. Otherwise, dyads should have modeled the closed-chain change problems as they had done for the open-chain change problems.

Anecdotal Evidence on Why Three Dyads Did Not Induce a Parity Rule

One dyad never discovered the locking outcome and sometimes continued for 300 s until the experimenter moved on to the next problem. The other two dyads who did not reach criterion solved all the problems but did not induce a parity rule. Instead, they found an alternative strategy to reduce their time and effort. Due to the $n + 1$ sequential presentation of the problems, they determined that the next problem would have the opposite answer of the immediately preceding problem. One dyad determined this by noting that the answers were alternating, and the other dyad simply added one more gear to each subsequent problem. For example, if the seventh gear of a seven gear chain turns clockwise, then the eighth gear in an eight gear chain should turn counter-clockwise because it touches a clockwise gear. Thus, two dyads induced procedures for solving the problems in their presented sequence instead of inducing a parity rule for solving the problems in any sequence. Their data looked similar to the data of the dyads who induced the parity rule. At first they modeled the problems in each configuration; latencies, rotational gestures and exophoric references were high. But once these two dyads induced the strategy of relying on the problem sequence, all the measures dropped. The number of quantitative references, however, did not increase. They were not concerned with the numerosity of a problem so much as they were concerned with the behavior of the last gear in the preceding problem.

Discussion

The explicit statements of the parity rule indicated that the dyads induced their rules around the third problem for each gear configuration. This supports the *a priori* separation of problems into blocks of three in Experiment 1. The changes that occurred in dyad behavior before and after rule statement fit our overall story. Prior to the statement of the rule in each configuration, dyads exhibited modeling behavior through long latencies, frequent rotational gestures, and exophoric references (e.g., "it goes *this* way"). After stating their rules, dyads exhibited rule-based behaviors through short latencies, few gestures, and frequent quantitative references.

The change problems provided evidence supporting two hypotheses regarding fallback modeling. The modeling for the open-chain change problems supports the hypothesis that people will generally model when their rules are too narrowly defined, and not just in the especially difficult case of the locking gears. For example, although dyads had a parity rule for the open-chain problems that began with a clockwise motion, they modeled for a new problem that began with a counter-clockwise motion. The second hypothesis was that the closed-chain problems are not treated as completely novel problems. Supporting this hypothesis, the lack of modeling for the closed-chain change problems indicates that dyads had induced a closed-chain rule that incorporated information from their open-chain solutions. The dyads had not solved a closed-chain problem with an initial counter-clockwise motion, and therefore, they must have mapped the information from the open-chain change problems that they had modeled.

The empirical evidence indicating when dyads modeled suggests a rational analysis that differentiates three modeling situations—novel, generalization, and rule-failure situations. Each of these situations is precipitated by the inadequacies in one's body of rules. However, one can differentiate the character of these inadequacies. Even though people may be blind to their type of rule inadequacy prior to developing surrogate evidence, the different situations have consequences for how much modeling will be necessary to develop a new rule.

To develop the analysis of the three modeling situations, consider a parity rule as having an input–output form. For example, if the input is odd then the output is clockwise. In a novel situation where people have no relevant rules, the range of the output is necessarily undefined; one does not know the range of possible behaviors available to a system. The initial open-chain modeling is an example. From debriefing, we found that participants were not previously familiar with problems of this structure. Thus, the depictive evidence of clockwise and counter-clockwise motions helped define the possible outcome space of clockwise and counter-clockwise. In a generalization situation, the output range of a rule is correct, but the input range is too narrow. The modeling for the open-chain change problems is the relevant example. At this point, the dyads' rules already had the outcome range of clockwise and counter-clockwise but did not have provisions for a counter-clockwise (or right-hand side) input for the initial gear. This led to different empirical outcomes than in the novel problem situation. In the novel situation, people took three trials to induce a rule. In the generalization situation, dyads only required one trial to refine their rules, as shown by the fact they did not need to model the change problems in the closed chain. Finally, in a fallback situation, an existing rule has an output range that does not include the correct possibility and people need evidence to find the missing outcome. Once found, people may capitalize on the knowledge in their original rule. Evidence supporting a unique category for fallback modeling comes from the closed-chain problems. First, unlike the generalization situation, dyads needed to discover the locking outcome and it took 2.7 successful trials to induce a new rule (although one locking problem is sufficient for inferring the rule modification). Second, dyads did not model the closed-chain change problems. Presumably, this was because their induced closed-chain rule incorporated the generalized rule developed for the open-chain change problems. Thus, the first closed-chain problems were not completely novel because dyads could use their open-chain knowledge to inform their closed chain rule.

To further refine the account of the interplay between models and rules, we consider why two dyads induced a problem-solving solution that we call the *Add-On* strategy. These dyads solved each successive problem by viewing it as an $n+1$ problem where the n represents the prior problem and the $+1$ represents the addition of one more gear in the current problem. An issue that may bear on the transition between models and rules is why these two dyads induced the Add-On strategy but not the parity rule. Our hypothesis is that the Add-On strategy inhibited the discovery of the parity rule because it made a gear's numerical position within a chain irrelevant. For example, one could treat a fourth gear as "one more," and consequently its evenness would not be readily available. Because the numeric positions of the gears were not simultaneously represented with their motions, the parallel alternation between motions and number parities may not have been detected. Experiment 3 examined this possibility.

EXPERIMENT 3

These first two experiments primarily focused on the conditions that precipitate depiction. The current experiment explores how people move from a depictive representation to a numerical one by examining the representations that mediate the transition. The first two experiments treated the transition from models to rules as though it were instantaneous. However, as pointed out by Duncker (1945) and others (e.g., Kaplan & Simon, 1990), there is a gradual restructuring of a problem prior to a moment of insight, inductive or otherwise. Prior to the moment of inductive insight, we saw hints of two restructuring processes. Each of these processes yielded a representation that set the stage for fashioning a rule that no longer required analog spatial information. We call the two aspects of the first process, *fading* and *codifying*. In fading, attributes of the referents are removed from the model. In codifying, the results of the model simulations are codified into verbal labels. In the second process, *quantitative casting*, the gear motions are corepresented with the numerical properties of the gears. It is our hypothesis that fading and codifying are natural outcomes of repetitious modeling, and that quantitative casting enables the move from a simple empirical pattern to a number-based rule. The following paragraphs delineate these processes in more detail. Figure 10 provides a schematic.

One of the strengths of depictive modeling is the lack of abstraction. It allows people to reason over a relatively complete representation of a situation before they have selected attributes for special attention (Reese, 1970). This may give depictive models some of their power in handling novel situations. They provide an opportunity to consider aspects of a problem that a symbolic rule might exclude (Schooler & Engstler-Schooler, 1990). By falling back to a model, for example, people can discover the jamming outcome that was not included in their open-chain parity rule. A model's inclusiveness, however, is also a liability. People are not able to "run" models of complex systems whole scale (Hegarty & Sims, 1994). Thus, when depictive models become too complex, or are repeatedly deployed, people need ways of simplifying them. Fading and codifying are two processes by which depictive models become simplified.

Fading occurs when elements of the original problem representation are dropped, or are replaced by a label through codification. For example, on first exposure to the problems people might imagine the texture of the gear faces (e.g., a smooth metallic sheen). As texture plays no role in the simulations, it fades quickly leaving a more schematic image of the gears. In contrast, people might also begin with a depiction that includes the nonslipping surfaces of the gears. This aspect of the model is essential for inferring the forces interacting between the gears. Consequently, this feature may not fade until the dynamics of the gears can be codified into a pattern of alternating motion labels.

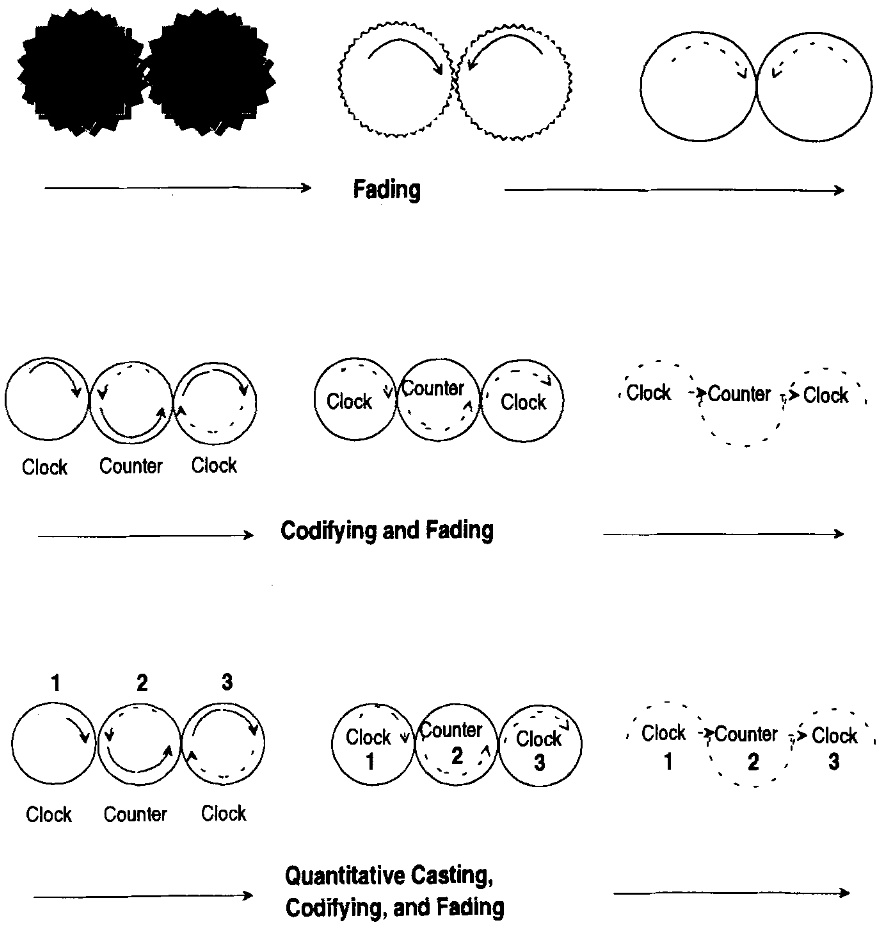


Figure 10. A Schematic Example of Fading, Codifying and Quantitative Casting.

Codifying refers to the process by which the causes and intermediate results in a simulation are labeled into discrete entities. For example, while propagating motion along the chain of modeled gears, people might state "clock, counter, clock..." Eventually, they might apply the pattern of alternating labels to static spatial tokens without a depiction of the gear motions. For example, they could imagine the gears as tick marks on a line, or use a separate finger to represent each gear in an external representation. They might say "clockwise" and lift one finger, then say "counter-clockwise" and lift the next finger, and so forth. By abridging the dynamic aspects of the model into a simple, labeled regularity, they no longer need to model the forces acting between the gears (cf. Weld, 1986). It is our hypothesis that people fade and codify the dynamics of their models over repeated trials.

Supporting evidence would be found if people switched from rotational hand gestures to "ticking" hand gestures by which they simply mark off an imaginary gear without representing its dynamics.

Fading and codifying diminish the demands of maintaining a continuous spatial transformation. One question is whether these faded representations still model nondynamic spatial properties of the gears, such as the left to right propagation of changes, or whether they are more like an ordered list (e.g., 1,2,3...). One way to address this question is to see what happens if a problem switches the position of the initial gear from the left- to right-hand side of the gear chain. If the codified representations still model spatial relations, then people should change the direction in which they "tick" through the gears.

To achieve the final abstraction in which spatial and temporal order are no longer modeled, quantitative casting may play an important role. Quantitative casting is a subset of symbolic casting in which the properties and relationships within a model are brought into the purview of a higher order and more articulated symbol system. Empirical pattern finding is only one piece of the induction story. For example, simply aggregating (Weld, 1986) or chunking (Laird, Rosenbloom, & Newell, 1986) the results of previous simulations is insufficient to transform these aggregations into a rule that relies on numerical properties like odd and even. A second piece of the story involves linking the empirical patterns to a symbol system with predefined structural properties, such as odd and even. This type of linking, or casting as we call it, occurs frequently in scientific research when a phenomenon is distilled to a few critical features, parameterized, and folded into the structure of a theory.

In the current case, quantitative casting may occur through the simultaneous consideration of number and model. As one enumerates, "Gear one is clockwise, gear two is counter-clockwise...," the alternating sequence of gear motions helps organize the numbers. In terms of our metaphor, the numbers are cast, or shaped, around the patterns generated by the model. This numerical shaping supports the realization of an alternating mathematical structure—odds and evens.

As an alternative to casting, one might consider trying to map numbers onto the gear problems whole scale. Imagine, for example, that a person has modeled short chains and is then confronted with a problem of 120 gears. Further imagine that at this point the reasoner has only abridged the model into a pattern of alternating labels. As a problem of this magnitude is cumbersome to model completely, the person may attempt to find a mathematical principle that can simplify the problem. However, one confronts the question of which mathematical principle to apply; a sequence of numbers has innumerable patterns. As a representation of number has not been molded to the alternating motion, there are few constraints on what pattern is chosen. For example, one could try a *factoring* strategy by looking for factors that divide evenly into 120. A salient number is ten. Subsequently, one might try to find a ten-gear rule, such that given the motion of the first gear, one can quickly determine the motion of the tenth gear. Finally, one might work through 12 applications of the ten-gear rule. Although this approach seems inefficient, if the pattern of odd and even has not been foreshadowed in the model, factoring is a salient approach for breaking down large numbers into more manageable chunks.

As a provisional test of the value of quantitative casting, we examined whether participants who corepresented the number and motion of each gear induced a parity rule. We identified this corepresentation through *Enumerations*. An enumeration occurs when people explicitly count out the number of a gear as they indicate its direction of motion, either gesturally or verbally; for example, "One is clockwise, two is counter-clockwise..." The effect of corepresenting number and motion through enumerations was compared to the effect of *Iterations*. With iterations, an individual moves through the chain of gears without explicitly counting each gear. If our hypothesis about the value of quantitative casting is correct, then people who use an enumerative strategy should induce parity rules more often than people who simply use an iterative strategy. An alternative hypothesis might be that people induce the parity rule solely on the basis of the association between the total number of gears and the final answer, regardless of the evidence generated by their modeling activity. If so, enumerations should not yield an inductive advantage.

In the first two experiments, the intermediate representations between rule and model passed too quickly to observe in detail. In this experiment, using 26 open-chain problems, two manipulations slowed down and stalled the process of induction so that we could see the component processes more clearly. One manipulation was to use numerically unsophisticated participants who would hopefully be more transparent than graduate students in their efforts to apply arithmetic concepts. The second method for capturing intermediate representations was to encourage participants to find a simplification to modeling, such as a parity rule, and then remove its advantages for subsequent problems. To encourage a simplification, there were problems with 121, 66, 218, and 51 gears. These problems are so large that they cue one to find a more efficient solution than modeling. However, before any resulting rule could become secure, the next problem changed the parameters of the problem (e.g., the direction of motion) or used a simple and previously solved problem. For the former case, the new rule may be ineffectively narrow for the changed problems. For the latter case, it may be more effortful to use the new rule than to replay a simulation for a simple problem. If these prevented rule use, participants would not have a chance to practice and make their rules firm. This would have the effect of putting the participants on the fence between their rules and models and allow observations of the representations that stand between the two.

Method

Participants
Sixteen 10th- and 11th-grade students from a boys parochial school in New York City voluntarily participated in the experiment. None of the boys had entered the later sequences of high school mathematics and were below average on standardized tests of mathematics and science achievement (Cooperative Admissions Examination, NEDT, and SAT). One boy was unable to complete the task. Two others stated, after the fact, that they already knew the parity rule, because they were slot car enthusiasts. These two participants unerringly and immediately solved all the problems without the use of gestures. Because the focus is on the intermediate representations, these participants were not included in the data analyses. Thus, we report data from 13 participants.

Design
There were 26 open-chain problems. Twenty-two problems had from 3 to 10 gears, and four had more than 50 gears. A small problem always followed a large problem to lure the student back into modeling. The base of Figure 12 provides an overview of the problem sequence and the problem categorization according to chain size and the initial gear's motion and position. Contiguous, small-chain problems that did not vary the position or motion of the initial gear were presented in either strict or loose ascension (e.g., 3,4,5,6 vs. 3,5,4,6). This manipulation had no effect and will not be considered further.

Procedure
Individual participants faced a video camera with the experimenter out of view. They were directed to be sure of their answer, and if they answered a problem incorrectly, they were to continue working on the problem until they reported the correct answer. The participants received instructions to think aloud. To practice these directives, the participants solved a warm-up problem about car steering. The experimenter then showed a two-inch spur gear and stated, "All of the following problems will be about gears that look like this one." Otherwise, the procedures followed Experiment 2.

Coding
A problem trial was coded as eliciting *rotating* motions, *ticking* motions, or *no motions* (i.e., neither of the other two). As before, rotating motions indicated that people were depicting the gear dynamics. Ticking motions indicated that people had faded the dynamics of their model but were still working sequentially along the gear chain. A ticking gesture occurred when a participant raised or pointed his fingers or hands (without rotating) for at least three times successively. No motion indicated that a participant was no

longer using an external model, as would be the case if he employed a parity rule. To simplify coding, and because rotating and ticking gestures rarely occurred during the same problem, each trial received a single coding. Of the total 338 trials across subjects, 8 had both ticking and rotational gestures. Those 8 are counted as half occurrences in each of the two gesture categories. An independent judge and the primary coder had 95% agreement on categorizations using a random sample of four individuals.

The problem-solving strategies were coded into mutually exclusive categories with each problem receiving a single coding. Although infrequent, if people used multiple strategies within a single problem, they were coded with their final strategy prior to answering the problem. The coding scheme was objective and only 4% of the problems received an *Unknown* categorization. The *Iterative* category was indicated either by a succession of alternating verbal labels (e.g., "clock, counter, clock"), or by a succession of three or more ticking or rotational hand movements. A problem fell into the *Enumerative* category if participants explicitly mentioned the number of each gear in conjunction with either a statement of the direction of that gear, or a ticking or rotational hand movement for that gear. The Iterative and Enumerative categories were mutually exclusive. The *Add-On* category was determined according to one of three criteria: (a) a participant explicitly stated that he was tacking on one or two more gears to the result of the previous problem; (b) the use of a single ticking or rotational motion (or, two if the new problem were larger than the last); or (c) the person responded quickly and correctly. To ensure that the fast response was not due to a parity rule, it had to be coupled with a slow response for the next problem that involved a large number of gears. For the large problems, the Add-On strategy is ineffective. Thus, if a person took a long time for the next large-chain problem, he did not know the parity rule on the prior problems and was using the Add-On strategy. The *Parity* category was determined by an explicit mention of "odd" or "even," or a short latency coupled with an explicit mention of "odd" or "even" on a prior or following problem with no intervening alternate categorizations. The *Factor* category was easily identifiable as the boys verbally searched for good factors, worked out divisions, and pondered remainders or fractional answers. Nobody considered the factor of two which might have lead to the discovery of the odd/even pattern. A *Guess* coding was used when a boy stated that he was guessing.

Although the cross-tabulation of the gesture and strategy codes in Table 3 does not bear upon the predictions in this experiment, these data help validate the use of the different categories of gesture and strategy. As in the case of the first two experiments, the parity rule was not accompanied by gestures whereas modeling was. This result is somewhat biased in that gesturing helped identify how a problem-solving trial should be categorized. It should also be noted that the relatively few Enumerative trials, compared to

TABLE 3
Frequency of Problems by Strategy and Gesture Types (Experiment 3)

Gesture	Iterate	Enumerate	Add-On	Parity	Factor	Guess	Unknown	Totals
			Strategy					
Rotating	30.5[a]	17.5	3.0	6.0	0.5	0.0	0.0	57.5
Ticking	75.5	7.5	3.0	1.0	7.5	1.0	0.0	95.5
Neither	23.0	5.0	26.0	90.0	5.0	19.0	14.0	185.0
Totals	129.0	33.0	32.0	97.0	13.0	20.0	14.0	338.0

[a] Trials on which a participant used both rotating and ticking gestures received a half coding in each category.

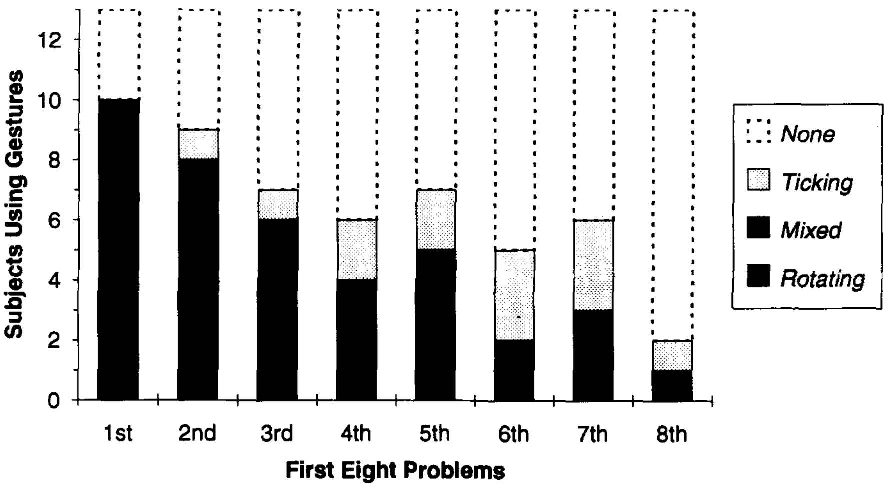


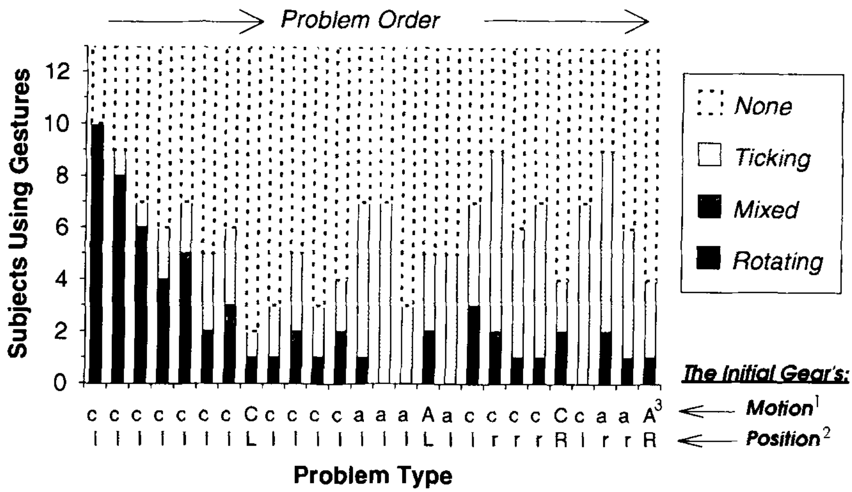
Figure 11. Changes in Gesturing over the First Eight Problems (Experiment 3).

Iterative trials, resulted from the fact that those who enumerated learned to use the parity rule (see below). In contrast, those who iterated did not find the parity rule and employed the Iterative strategy across problems.

Results

Gestural Evidence for the Progression to Faded Models

The first analysis considers the fading and coding of the models. Figure 11 shows the number of participants who solved the problems with rotating gestures, ticking gestures, both gestures, or neither for the first block of 8 problems. There was a progression from rotating gestures to ticking gestures to neither for the largest problem of 131 gears (i.e., problem number 8). These results reflect the progression from (1) a dynamic gestural model to



[1] C = clockwise; A = anti-clockwise motion.

[2] L = left side; R = right side of gear chain.

[3] Captial letters indicate the 4 large gear chains.

Figure 12. Changes in Gesturing over all Twenty-six Problems (Experiment 3).

(2) a faded static model that attached alternating labels to simple ticking motions to (3) a numerical rule that replaced the labels and static model altogether.

Through the full sequence of 26 problems shown in Figure 12, over 50% of the rotating gestures occurred on the first six problems compared to 10% of all the ticking motions. Twelve participants used rotating gestures during the experiment, of which nine also used ticking gestures. Once these nine participants progressed to ticking motions, they rarely returned to their more dynamic, rotating motions. For each problem that these nine participants solved using ticking gestures, the average number of previous problems that used rotating gestures was 4.7, whereas the number of following problems that used rotating gestures was 0.3; $F(1,8) = 16.3$, $MSe = 5.2$, $p < .01$. This indicates that the gestural models that represented rotational motion were faded into models that simply represented the position of each gear through a simple ticking motion.

The next analysis investigates whether the faded and codified models included spatial information. By the nineteenth problem, which first switched the initial gear from the left- to the right-hand side, one participant was using rotational gestures and eight were using the ticking gestures that indicate they had faded the gear dynamics. These eight participants are of interest because they can test whether the faded representations still modeled spatial information. If they did, then changing the location of the initial gear should have affected individuals' reasoning over their models. If the faded models were simply ordered lists (e.g., 1,2,3...), the location change should have little effect. When the problems switched the initial gear from the left- to right-hand side of the chain, seven of the eight participants reversed the direction of their ticking motions. For all 90 prior problems during which these eight participants had modeled the gears with their hands, they had proceeded left to right, but then on the first right-hand problem, seven of the eight participants reversed and moved right to left. During this reversal they used ticking gestures.

Verbal Evidence for the Effects of Quantitative Casting on Parity Rule Induction

The following analyses consider whether the simultaneous consideration of gear number and motion facilitated rule induction, as would be predicted if quantitative casting is an important piece of the induction process. The effect can be seen most simply by partitioning the individuals according to the relative frequency of enumeration. An Enumerative trial indicates that a participant was corepresenting number and motion (e.g., one-clock, two-counter...). Dividing the number of Enumerative trials by the total number of Enumerative and Iterative trials yields the percentage of modeling trials that were Enumerative.[2] No participant had an enumeration rate between 6% and 15%. This gap provided a natural partition of high and low users of enumeration. Table 4 shows that those participants who frequently enumerated also induced the parity rule; $\chi^2(1) = 4.99, p < .05$. For those who discovered the rule, the average percentage of modeling trials that were categorized as Enumerative was 34.2% (34% *SD*). In contrast, for those participants who did not discover a parity rule, the average percentage of Enumerative trials was 3.0% (3% *SD*).[3]

An alternative analysis using strategy frequency also shows the importance of quantitative casting. The number of Enumerative and Iterative problems for each person were within-subject measures capturing the factor of problem-solving strategy. Whether a person induced a parity rule

[2] We do not include the Add-on problems. There was no way of determining how people figure how many gears to add to the prior chain. For example, if the previous problem had 7 gears and the new problem had 9 gears, we could not tell if the participant was silently counting, "7,8,9," or if the participant silently thought, "I need to model two more."

[3] Of interest are the two participants who did not enumerate but did induce a parity rule. One had no modeling episodes at all, making it impossible to calculate a percentage. The other had six Iterative trials and no Enumerative trials. Two possible explanations for how these two participants managed to induce the parity rule without explicitly enumerating are that either they silently enumerated, or they detected the odd-even pattern by using the motion of the final gear and the total number of gears in the chain.

TABLE 4
Frequency of Participants Who Induced Parity Rule
Broken Down by High and Low Users of Enumerations (Experiment 3)

Percent of Modeling Episodes Coded as Enumerative	Induced Rule	Never Induced Rule
Participants with less than 6%	2[a]	5
Participants with more than 15%	6	0

[a] One of these two participants had no Iterative or Enumerative episodes.

was a between-subjects factor. For those eight individuals who induced a parity rule, the average number of problems using an Enumerative strategy was 3.8 and the number of problems using an Iterative strategy was 4.6. For those five individuals who did not induce a parity rule, there were 0.6 Enumerative problems and 18.4 Iterative problems. This created a reliable interaction between strategy and whether a person induced a parity rule; $F(1,11) = 20.47$, $MSe = 21.5$, $p < .01$.

Of the five individuals who did not learn the parity rule, four used a factor solution. These four boys spent considerable time and effort looking for factors that divided evenly into the total number of gears. Some of the factors attempted were 3, 7, 10, and 11. The fifth boy indicated that he was guessing by stating that he always used the opposite of the first gear's direction for large problems. In these five cases, the factor and guessing solutions were only applied to problems with large numbers of gears. On the other problems, these children primarily used either the Add-On or Iterative strategy.

Discussion

The evolution of participants' gestures and strategies provides evidence on the intermediate representations that bridge between depictive and numerical representations. Generally, participants first gesturally modeled the full motions of the gears and stated the direction of rotation for each hand. Next, the dynamics of the model were replaced with a simple method of successively alternating between the opposites of clockwise and counter-clockwise. At this point, a spatial model was still in play even though the participants had faded their gestures into a simple ticking motion and had codified the gears' alternating motions into labels. When the problem started on the right-hand side, rather than the left, participants changed the direction that they ticked the gears off on their fingers. Simultaneous with the fading and codifying, some individuals cast quantitative properties onto each gear. Specifically, as they noted the direction of a gear, they also stated the number indicating the ordinal position of the gear. This casting of numbers was important for the arithmetic properties of odd and even to arise with the pattern of motion alternation. Although the final form of the rule was based on the total number of gears and the final answer, shaping the numbers to the quasi-empirical structure of the model facilitated the induction of the global rule (cf. Kieras & Bovair, 1984). The five boys who did not induce the parity rule did not numerically label each gear. Consequently, when the problems involved numerous gears, they had few constraints on their mathematical reasoning and searched wildly for factors.

Although the foregoing progression is a fair portrayal of the sequence of events, the measures of modeling were fairly coarse. For example, one cannot be sure that when the participants switched from the rotating to ticking gestures that they did not simply substitute internal imagery of rotating gears for the external gestures. Moreover, there was no control condition. Consequently, our causal attributions are speculative. For example, one cannot be sure that quantitative casting enabled the induction of the parity rule. Perhaps, rather than mediating the discovery of the rule, the enumerations were simply a side-effect of some underlying variable that also yielded the parity rule. To test the functional role of quantitative casting one would need a more intrusive methodology than employed here. So, just as one might ask people to sit on their hands to see if gestures play a functional role in solving the gear problems, one might also prevent people from enumerating as they solve these problems.

GENERAL DISCUSSION

Formulating depictive models as surrogates for perceptual evidence generated a hypothesis space that was borne out through over-lapping predictions and convergent lines of evidence. As in situations where people would want to observe the world, modeling occurred when people confronted a novel problem, when people needed to generalize rules, and when their rules failed. In these three situations, people modeled the problems on their hands, took a relatively long time conducting their simulations, and used language referring to their models.

Like observations of nature, model simulations served as the raw data for inducing abstract rules. This occurred when people abstracted regularities from their models' behaviors and cast numerical concepts about these regularities. Changes indicating the induction of a parity rule occurred around the third problem as predicted on the theoretical grounds that it takes at least two examples to induce a binary pattern. Empirically, people actually stated a parity rule at the point where the behaviors changed. Figure 10 provides a schematic of how people transformed a dynamic model into a static model with alternating motion labels. After their initial simulations, people no longer modeled the dynamic relations between the gears using rotating gestures. Instead, they codified the results of the simulations into a pattern of alternation by which they could subsequently point to an imaginary gear and state its direction of motion. This process is illustrated in the second row of Figure 10. The third row of the figure represents the casting of numbers onto the model depictions. The importance of casting an articulated symbol system onto the model was demonstrated by the fact that those individuals who did not induce a parity rule were also the ones who did not cast quantities onto their models.

To complement the conceptualization of depictive models as surrogates for experience, one may think of numerical rules as simple theories about experience. Reflecting the properties of theories, many participants employed rules when they had them. The rules formalized attributes of the problems and did not require the time and effort of multiple simulations chained together. Moreover, the rules allowed individuals to reason about symbolic relationships like contradiction and parity. Quick response times provided evidence of rule use, the disappearance of rotating gestures provided evidence of reasoning without a dynamic model, and the switch to quantitative language provided evidence of reasoning that relied on numerical relationships. The error data from Experiments 1 and 2 suggest that people tended to over-extend their rules before they fell back to modeling, much as one might over-extend a theory prior to its disconfirmation. The bulk of errors occurred when the parameters of a problem changed in such a way that a newly induced rule would not apply. However, many of these errors may be simply attributable to the increased difficulty of the closed-chain problems. Further research is necessary to develop a clearer understanding of the tendency to misapply a rule (or, theory) when the evidence from a model (or, experiment; Kuhn, 1989) would be more appropriate.

In general, depictive models are useful in situations where one does not have a formal method for deriving outcomes; they generate phenomena unknown to a limited collection of rules. Rules are useful in that they are efficient and can use formal properties like contradiction and oddness. Their separate strengths are perhaps best captured by one boy in the third experiment. Although he induced a parity rule after the first few problems, the rule was too narrow and ignored important problem parameters. His rule could be stated, "If the problem has an odd number of gears, then the last gear turns clockwise." When confronted with a new problem in which the initial gear turned counter-clockwise, he applied his previously induced parity rule and reached the wrong answer. Rather than generalizing his parity rule to handle the parameter change, he settled on a compromise between the speed of the rule and the evidence of the model. He created a rule stating that all the even gears turn the same way, as do the odd gears, and he modeled to see which way the first odd and even gears were turning. For example, he would first note the parity of the problem as being even. Next he would model two gears to determine which way the first even gear in the chain was turning. Because of his rule, he could conclude that the motion of this even gear would be the same as the motion of the final gear which was also an even. Given the lack of a more general parity rule, this is an ingenious solution that gets the best of both the modeling and analytic worlds. We suspect that this strategic shuttling between depictive evidence and symbolic formulation is highly characteristic of reasoning about physical systems.

REFERENCES

Alibali, M.W., & Goldin-Meadow, S. (1993). Gesture-speech mismatch and mechanisms of learning: What the hands reveal about a child's state of mind. *Cognitive Psychology, 25,* 468–523.

Black, J.B., & Bower, G.H. (1980). Story understanding as problem solving. *Poetics, 9,* 223–250.

Bransford, J.D., Barclay, J.R., & Franks, J.J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology, 2,* 193–209.

Bruner, J.S. (1966). On cognitive growth. In J.S. Bruner, R.R. Oliver, & P.M. Greenfield (Eds.), *Studies in cognitive growth.* New York: Wiley.

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

Clement, J. (1994). Use of physical intuition and imagistic simulation in expert problem solving. In D. Tirosh (Ed.), *Implicit and explicit knowledge.* Hillsdale, NJ: Erlbaum.

Cliff, N. (1987). *Analyzing multivariate data.* New York: Harcourt Brace Jovanovich.

Collins, A., Brown, J.S., & Larkin, K.M. (1980). Inference in text understanding. In R.J. Spiro, B.C. Bruce, & W.F. Brewer (Eds.), *Theoretical issues in reading comprehension: Perspectives from cognitive psychology, linguistics, artificial intelligence, and education.* Hillsdale, NJ: Erlbaum.

de Kleer, J., & Brown, J.S. (1981). Mental models of physical mechanisms and their acquisition. In J.R. Anderson (Ed.), *Cognitive skills and their acquisition.* Hillsdale, NJ: Erlbaum.

de Kleer, J., & Brown, J.S. (1983). Assumptions and ambiguities in mechanistic mental models. In D. Gentner & A.L. Stevens (Eds.), *Mental models.* Hillsdale, NJ: Erlbaum.

diSessa, A.A. (1988). Knowledge in pieces. In G. Forman & P. Pufall (Eds.), *Constuctivism in the computer age.* Hillsdale, NJ: Erlbaum.

diSessa, A.A. (1993). Toward an epistemology of physics. *Cognition and Instruction, 10,* 105–225.

Duncker, K. (1945). On problem solving. *Psychological Monographs, 58* (Whole No. 5).

Ekman, P., & Friesen, W. V. (1972). Hand movement. *The Journal of Communication, 22,* 353–374.

Forbus, K. (1990). Qualitative physics: Past, present, and future. In D.S. Weld & J. de Kleer (Eds.), *Readings in qualitative reasoning about physical systems.* San Mateo, CA: Morgan Kaufmann.

Funt, B. V. (1980). Problem solving with diagrammatic representations. *Artificial Intelligence, 13,* 201–230.

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

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

Halliday, M.A.K., & Hasan, R. (1976). *Cohesion in English.* London: Longman Group Limited.

Hegarty, M. (1992). Mental animation: Inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Learning, Memory & Cognition, 18,* 1084–1102.

Hegarty, M., & Just, M.A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language, 32,* 717–742.

Hegarty, M., Just, M.A., & Morrison, I.R. (1988). Mental models of mechanical systems: Individual differences in qualitative and quantitative reasoning. *Cognitive Psychology, 20,* 191–236.

Hegarty, M., & Sims, V.K. (1994). Individual differences in mental animation during mechanical reasoning. *Memory & Cognition, 22,* 411–430.

Hendrix, G.G. (1973). Modeling simultaneous actions and continuous processes. *Artificial Intelligence, 4,* 145–180.

Holland, J.H., Holyoak, K.J., Nisbett, R.E., & Thagard, P.R. (1986). *Induction: Processes of inference, learning & discovery.* Cambridge, MA: The MIT Press.

Johnson-Laird, P.N. (1981). Mental models of meaning. In A.K. Joshi, B.L. Webber, & I.A. Sag (Eds.), *Elements of discourse understanding.* Cambridge: Cambridge University Press.

Johnson-Laird, P.N. (1983). *Mental models.* Cambridge, MA: Harvard University Press.

Kaplan, C.A., & Simon, H. (1990). In search of insight. *Cognitive Psychology, 22,* 374–419.

Kieras, D.E., & Bovair, S. (1984). The role of a mental model in learning to operate a device. *Cognitive Science, 8,* 255–273.

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12,* 1–48.

Klatzky, R.L., Pellegrino, J.W., McCloskey, B.P., & Doherty, S. (1989). Can you squeeze a tomato? The role of motor representations in semantic sensibility judgments. *Journal of Memory and Language, 15,* 56–77.

Kosslyn, S.M. (1980). *Image and mind.* Cambridge, MA: Harvard University Press.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review, 96,* 674–689.

Laird, J., Rosenbloom, P., & Newell, A. (1986). *Universal subgoaling and chunking.* Boston, MA: Kluwer Academic Publishers.

Larkin, J.H., McDermott, J., Simon, D.P., & Simon, H.A. (1980). Expert and novice performance in solving physics problems. *Science, 208,* 1335–1342.

Laughlin, P.R., & Shippy, T.A. (1983). Collective induction. *Journal of Personality and Social Psychology, 45,* 94–100.

Mani, K., & Johnson-Laird, P.N. (1982). The mental representation of spatial descriptions. *Memory & Cognition, 10,* 181–187.

McNamara, T.P., Miller, D.L., & Bransford, J.D. (1991). Mental models and reading comprehension. In R. Barr, M.L. Kamil, P.B. Mosenthal, & P.D. Pearson (Eds.), *Handbook of reading research, (Vol. 2).* New York: Longman.

McNeil, D. (1987). *Psycholinguistics: A new approach.* New York: Harper & Row Publishers.

Metz, K.E. (1985). The development of children's problem solving in a gears task: A problem space perspective. *Cognitive Science, 9,* 431–471.

Parsons, L.M. (1994). Temporal and kinematic properties of motor behavior reflected in mentally simulated action. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 709–730.

Reese, H.W. (1970). Imagery and contextual meaning. *Psychological Bulletin, 73,* 404–414.

Resnick, L.B. (1987). Learning in school and out. *Educational Researcher, 16,* 13–20.

Rieser, J.J., Garing, A. E., & Young, M.E. (1994). Imagery, action, and young children's spatial orientation: It's not being there that counts, it's what one has in mind. *Child Development, 65,* 1254–1270.

Rieser, J.J., & Schwartz, D.L. (in preparation). *Hands in the mind's eye: The coupling of representation and action in imagined object rotations.*

Saltz, E., & Donnenwerth-Nolan, S. (1981). Does motoric imagery facilitate memory for sentences? A selective interference test. *Journal of Verbal Learning and Verbal Behavior, 20,* 322–332.

Schooler, J.W., & Engstler-Schooler, T.Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology, 17,* 36–71.

Schwartz, D.L. (1995a). The emergence of abstract representations in dyad problem solving. *The Journal of the Learning Sciences, 4,* 321–354.

Schwartz, D.L. (1995b). Reasoning about the referent of a picture versus reasoning about the picture as the referent: An effect of visual realism. *Memory & Cognition, 23,* 709–722.

Schwartz, D.L., & Black, T. (1996). *Situated physical inference: Depictive knowing versus descriptive beliefs.* Manuscript submitted for publication.

Schwartz, D.L., & Black, J.B. (1996). Analog imagery in mental model reasoning: Depictive models. *Cognitive Psychology, 30,* 154–219.

Scribner, S. (1984). Studying working intelligence. In B. Rogoff & J. Lave (Eds.), *Everyday cognition: Its development in social context.* Cambridge, MA: Harvard University Press.

Shepard, R.N. (1994). Perceptual-cognitive universals as reflections of the world. *Psychonomic Bulletin and Review, 1,* 2–28.

Shepard, R.N., & Cooper, L.A. (1982). *Mental images and their transformations.* Cambridge, MA: The MIT Press.

Simon, H.A., & Lea, G. (1974). Problem solving and rule induction: A unified view. In L.W. Gregg (Ed.), *Knowledge and cognition.* Potomac, MD: Erlbaum.

Van Dijk, T., & Kintsch, W. (1983). *Strategies of discourse comprehension.* New York: Academic.

Vosniadou, S., & Brewer, W.F. (1993). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology, 24,* 534–585.

Weld, D.S. (1986). The use of aggregation in causal simulation. *Artificial Intelligence, 30,* 1–17.

White, B., & Frederiksen, J. (1990). Causal model progressions as a foundation for intelligent learning environments. *Artificial Intelligence, 24,* 99–157.