

Using Scientific Inquiry Activities in Exhibit Explanations

SUE ALLEN

The Exploratorium, 3601 Lyon Street, San Francisco, CA 94123, USA;

e-mail: sueallen@research.com

Received 5 November 1996; accepted 15 May 1997

ABSTRACT: This study investigated the effect of different scientific inquiry activities on visitors' understanding of the science underlying an interactive exhibit. The exhibit, "colored shadows," creates a pattern of colored shadows on a white wall, due to a person's body blocking the light from colored lamps. The subjects were 392 museum visitors, aged 7 to adult. They were individually guided through a structured interview, during which they did one of seven inquiry activities, randomly assigned. The activities were: generate an explanation; interpret an explanation; troubleshoot an explanation; choose between two explanations; choose plus design a discriminating experiment; choose plus make and test a related prediction; and make a prediction before experiencing the phenomenon. As a test of their final understanding, visitors were asked to complete two near-transfer tasks in diagrammatic form. The results showed that the interpretation activity was the most effective in facilitating visitors' understanding of the mechanism of shadow-creation; least effective was the activity in which visitors made a prediction before experiencing the phenomenon. Visitors had relatively little difficulty choosing correctly from two explanations, but had much more difficulty designing a discriminating experiment. Only rarely during the interviews did visitors revise their thinking in the face of disconfirming evidence. Although the majority of visitors did not answer both transfer problems correctly, their thinking generally showed consistency and logic. The most common types of reasoning used were causal models that traced the path of light from each lamp to the object and to the wall. Such models incorporated ideas of reflection, blocking, and reflection/blocking hybrids, among others. A follow-up study will incorporate these insights into stand-alone exhibit labels for further testing. © 1997 John Wiley & Sons, Inc. *Sci Ed* **81**:715–734, 1997.

INTRODUCTION

At the Exploratorium and many other hands-on science museums, most exhibits are accompanied by explanatory labels. These may include text, drawings, and/or photographs, and are designed to help visitors understand what is going on behind a surprising or intriguing phenomenon. As yet, however, there has been no systematic account of how visitors make sense of these explanations: how visitors incorporate, reject, ignore, or modify them. This aspect of

A shorter version of this report was originally presented at the annual meeting of the Visitor Studies Association, Estes Park, CO, 1996.

Correspondence to: S. Allen

Contract grant sponsor: James S. McDonnell Foundation

cognition is particularly important in the domain of science, because it is well known that people come to the hands-on experience with robust pre-existing models of how the physical world operates (e.g., Arons, 1990; Pfundt & Duit, 1991; Hein, 1995), so they are not simply “empty vessels waiting to be filled with knowledge.”

To improve science learning in formal as well as informal settings, we need to know how people respond to surprising or contradictory phenomena, and how they use new information to build and refine their own knowledge structures. This view gives central importance to *scientific inquiry*, the loose aggregation of strategies that scientists use to advance their understanding of the complex world. Many educational researchers argue that such inquiry skills, emphasizing conceptual understanding and modeling, are more important for learning science than the acquisition of domain-specific facts and information (e.g., Brown & Campione, 1990; Burbules & Linn, 1991; Halloun & Hestenes, 1987; White, 1993). The same emphasis on scientific methods and critical thinking now appears in the Science Framework for California Public Schools (Science Curriculum Framework and Criteria Committee, 1990) and the *National Science Education Standards* (National Research Council, 1996).

RESEARCH QUESTION

This study is the first in a series that explores the effects of incorporating scientific inquiry into hands-on museum exhibits. The following candidates were considered inquiry activities that might be suitable to include in a hands-on exhibit: generating an explanation; interpreting a given explanation; troubleshooting an incorrect explanation; evaluating competing explanations; making predictions based on a model; designing an experiment to discriminate between models; and interpreting the results of a discriminating experiment.

In this study, I ask:

- Are all of these inquiry activities equally effective in facilitating visitors’ understanding of the science behind an intriguing phenomenon, or are some more effective than others? In other words, if visitors were to do these specific inquiry activities while at an exhibit, which one(s) would lead to the best understanding?
- How challenging are these activities to visitors? Are they all achievable?
- What is the nature of visitors’ understanding of the science behind a phenomenon after they have gone through such inquiry activities?
- How likely are visitors to revise their faulty mental models of a phenomenon if an inquiry activity leads them to face disconfirming evidence?

One assumption behind the study is that it is highly desirable that visitors understand the science behind an intriguing phenomenon. Of course, this is not the only or even necessarily primary goal of hands-on exhibits; yet it does represent a significant part of our educational goal at the Exploratorium. Another assumption is that exhibit designers are able to design exhibits that will lead visitors to undertake these inquiry activities spontaneously. This is not a simple assumption, and will be addressed in a follow-up study to this one. A third assumption is that studying one exhibit in depth can lead to insights that apply to a larger group of exhibits; this remains to be tested.

Because the main purpose of this study was to compare the effects on learning of different inquiry activities, visitors were constrained in certain ways. Specifically:

- (a) Visitors were asked for an initial time commitment of “5–10 minutes” so that they would have time to think carefully about the exhibit.

- (b) Visitors' attention was focused on one intriguing aspect of the exhibit, rather than encouraged to wander across a range of different intriguing phenomena with different scientific explanations.
- (c) Visitors were led through an interview at the exhibit, guided to do the inquiry activity we were testing, and no others.

Each of these constraints helped to standardize the experiment, providing "best case" information relevant to the research question posed, and giving a rich and detailed picture of each visitor's reasoning; on the other hand, each constraint also made the visitors' experience more removed from the unstructured, open-ended, social, spontaneous experience that visitors have when no interviewer is present. For this reason, the follow-up study to this one (in which the inquiry activities will be incorporated into stand-alone exhibits and visitors will be observed interacting in their social groups) will be an important complement.

METHOD

For reasons of simplicity, this study focuses entirely on individual visitor cognition. I conducted interviews with individual visitors at an exhibit, leading them through a randomly assigned inquiry activity and a final assessment question. The visitor's experience was thus mediated by the interviewer to ensure standardization of the questions asked and the resources made available.

Choice of Exhibit

I chose to study scientific inquiry activities in the context of the exhibit known as "colored shadows." In this exhibit, three lamps (red, green, and blue) shine onto a white wall from widely separated locations. When visitors move in front of the lamps, they generate overlapping colored shadows on the wall, because the blocking of light from any one of the lamps leaves colored light on the wall from the remaining two.

I chose this exhibit for the following reasons:

- The experience is readily accessible, without requiring visitors to follow a sequence of instructions.
- The physics that underlies the exhibit is not highly technical, and builds on intuitions about shadows that are familiar from everyday life.
- The exhibit is sufficiently open-ended to support a range of inquiry activities, including experimental design.
- The exhibit has been previously studied by researchers, and shown to evoke a misconception about the behavior of light from the lamps. Specifically, Feher (1990) and Feher and Rice (1986) originally reported that many visitors think the shadows are created by light bouncing off, rather than being blocked by, the object in front of the lamps. Similar findings were reported by Perry (1989) and Pfeiffer (1995). This documented misconception provided an excellent explanation to offer as an alternative in the inquiry activities involving choice.

Design of Interview

Interviews were used to test the effectiveness of different inquiry activities. The general structure of each interview was: (a) focus the visitor's attention on a particular intriguing

phenomenon; (b) ask the visitor to make sense of the phenomenon by doing a specific inquiry activity; and (c) assess the visitor's understanding of the exhibit following the inquiry activity. Throughout the interview, I recorded visitors' responses as accurately as possible on a notepad.

The details of the interviews were as follows: I began by inviting the visitor to move around and experience the beauty of the colored shadows. I then explained that I was going to switch off one of the lamps, so as to make the situation less complicated to understand. I switched off the green lamp in the center, leaving only the red and blue. The result was that the visitor cast two colored shadows on the wall: a red one on the same side as the red lamp, and a blue one on the same side as the blue lamp (see Fig. 1). I then directed the visitor's attention to the relationship between color and location of each shadow:

You can see you just have two shadows now: a red and a blue. Well, the first time I did this, what I found really strange was the way around the shadows are. I thought the blue lamp over there would make a blue shadow on the wall over here [point to the place diagonally across from the blue lamp], but it doesn't; the colors are the other way around. Does that seem strange to you as well? Or does it make sense to you the way it is?

The wording of this part of the interview was carefully chosen with two purposes in mind. Primarily, it was designed to focus visitors' attention on the specific features of the exhibit that were central to our explanatory model (viz., the precise arrangement and color of the shadows), by presenting plausible reasoning that might suggest a different result as a more

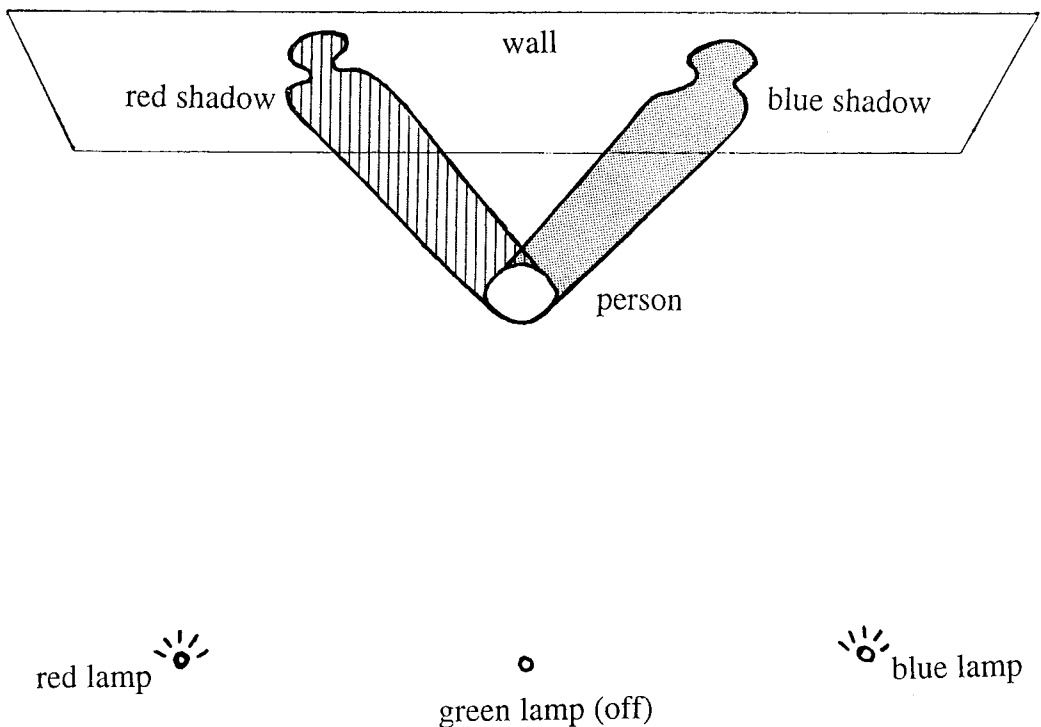


Figure 1. The intriguing phenomenon: A red shadow on the left, a blue shadow on the right.

likely outcome. In other words, the quoted reasoning served to emphasize that there really was something here that warranted explanation; without it, visitors might more easily have dismissed the complex shadow arrangement with: “Why not? They’re colored lamps, so they make colored shadows.” The statement was thus an attempt to standardize the visitors’ initial orientation with respect to the dilemma we were highlighting: (a) by showing exactly which dilemma we were asking them to grapple with; and (b) by showing that it was a genuine dilemma requiring an explanation.¹ Secondly, the use of the apparently flawed reasoning, followed by the questions, invited visitors to give the correct reasoning if they knew it. Thus, the speech had the potential to serve as a very rough assessment tool of visitors’ current understanding: if visitors already understood and recognized the blocking principles behind the exhibit, then the last question might invite them to say something about those principles.²

At this point, if the visitor simply responded that the exhibit did make sense, I probed for some further explanation. This was an attempt to distinguish visitors who had a basically correct understanding from those with a misconception. If the visitor responded that the shadow location did indeed seem strange, I simply agreed. In all cases, I then told visitors that my chief interest lay in this aspect of the exhibit (viz., accounting for the location and color of the two shadows).

I then asked the person to do one of seven inquiry activities, randomly assigned. The activities made use of different explanations which were given to visitors on cardboard, with text and diagrams, to simulate elements of a label. The activities were as follows:

- (A) *Generate an explanation* of the phenomenon, with no label provided. In this condition, I gave the visitor a schematic picture of the wall and lamps, and asked him or her to draw in the location of the shadows and to explain what the light was doing to cause them.
- (B) *Interpret the standard scientific explanation*, as presented in a single label. This corresponds to the control of the experiment, being the inquiry activity we most commonly ask visitors to do. In this condition, I provided the visitor with an explanation (including text and a diagram) of the fundamental mechanism of blocking and how it gave rise to the two shadows (see Fig. 2). I told the visitor that this was the explanation of the shadows, and asked him to tell me in his own words what it was trying to say, and whether it made sense.
- (C) *Troubleshoot an explanation*, as presented in a single label that makes an obviously wrong prediction. In this condition, I provided the visitor with an explanation which correctly showed the “crossing over” geometry of the two shadows cast on the wall, but omitted the idea of blocking by the person’s body (see Fig. 3). I pointed out that this explanation was flawed, in that it showed two shadows that were the wrong way around in terms of color. I asked the visitor to find the flaw in the explanation, and to tell me how fixing the flaw might account for the actual shadows’ locations and colors.
- (D) *Choose between two explanations* that cannot both be right. In this condition, I

¹The particular line of reasoning was chosen so as to minimize the possibility of introducing a new misconception about the mechanism of shadow formation. In fact, there is nothing wrong with the “crossing-over” *geometry* of shadow formation that the interviewer gesticulated. What is missing is the critical notion of blocking as a mechanism to explain why the *color* of shadow cast by the blue lamp is dark rather than blue.

²Of course, such an assessment tool is not ideal: visitors may in fact understand the mechanism of light blocking in shadow formation, and may simply not see it as salient in this situation. However, a full pretest was not feasible given time constraints.

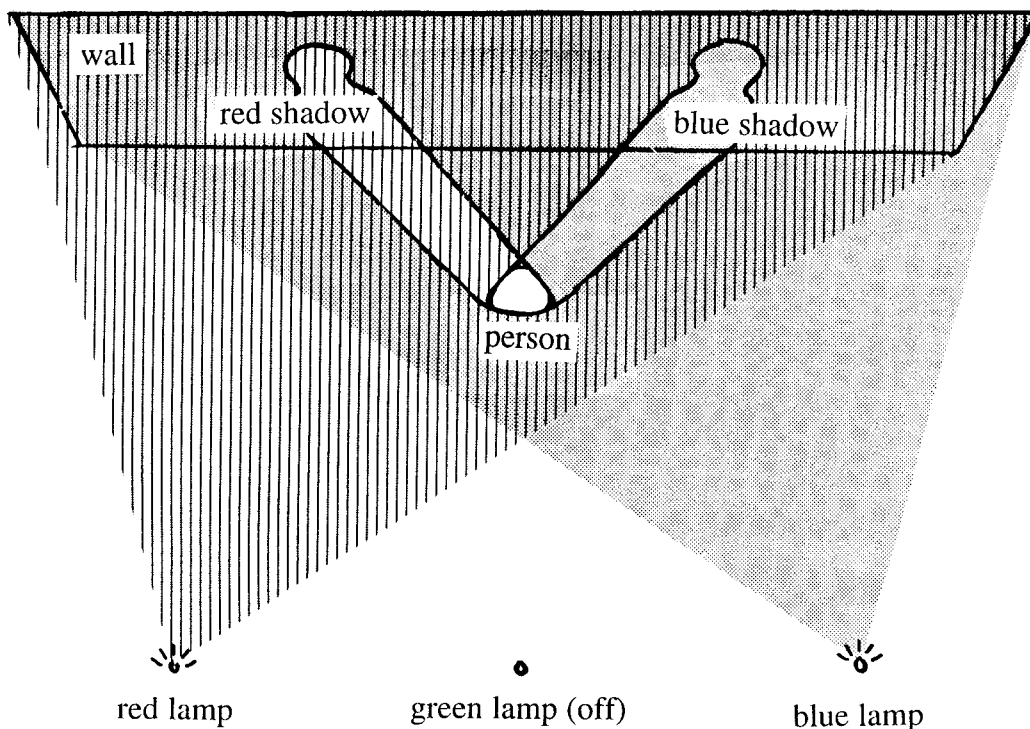


Figure 2. The scientific explanation (blocking): Both lamps shine over the whole wall. Where light from one lamp is blocked, the other color remains.

asked the visitor to choose which of two explanations was better, and to tell me why. One explanation accounted for the formation of the shadows in terms of blocking (the correct model, shown in Fig. 2), whereas the other used reflecting (the misconception, shown in Fig. 4).

- (E) *Choose between two explanations plus design a discriminating experiment.* In this condition, I began (as in condition D) by asking the visitor to choose between explanations based on blocking and reflecting. I then asked whether there was anything we could do or try that would prove which one was correct (i.e., discriminate between them). If the visitor was able to think of an idea that was feasible to do within the setting, then we carried it out and I asked him or her to interpret the result.
- (F) *Choose between two explanations plus make a related prediction.* This condition also began with my asking the visitor to choose between the two explanations of condition D. In this case, however, the follow-up activity was a prediction: I asked him or her to predict what would happen if we turned off the blue lamp, leaving only the red lamp on. Once the visitor had made a prediction, I switched off the blue lamp and invited him or her to interpret the result.
- (G) *Make a prediction before experiencing the phenomenon.* In this condition, I did not switch off the center lamp until the visitor had made a prediction of what would happen. To make this condition comparably difficult with the choice-based conditions, I gave the visitor two possible explanations which led to different predictions of shadow color, and I asked him or her to choose one. The

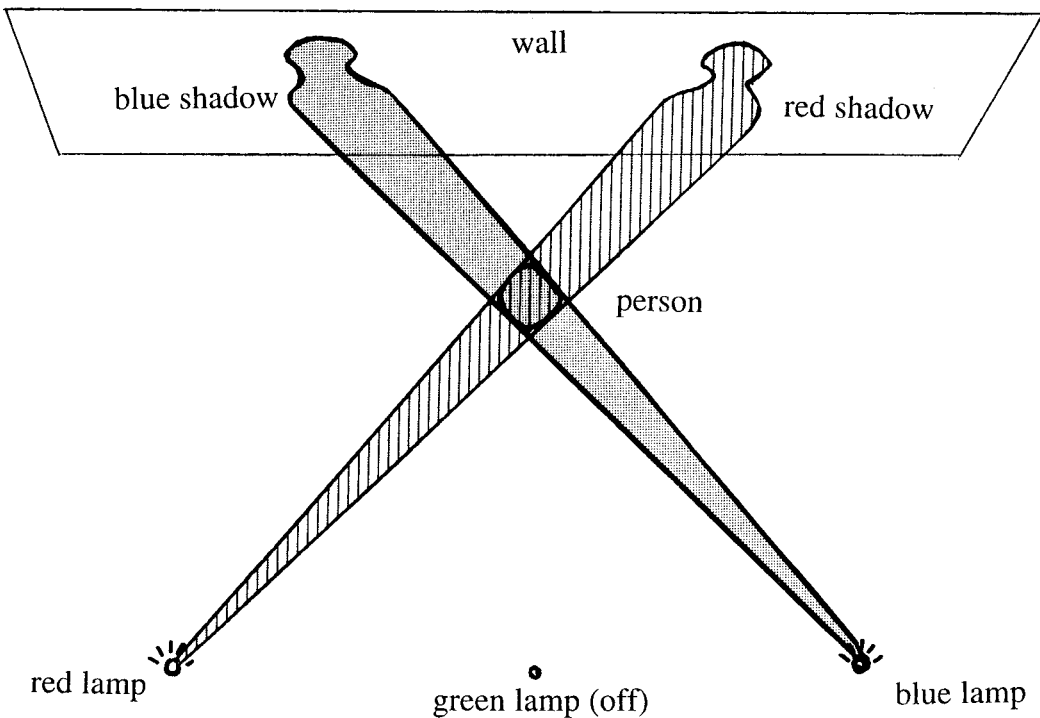


Figure 3. An obviously flawed model (crossing over): Shadows lead away from the light source, so the lamps throw two shadows that cross over. In reality, the shadows on the wall are oppositely colored.

explanations were based on blocking (Fig. 2) and crossing over (Fig. 3). I then switched off the lamp and invited the visitor to interpret the result.

Design of Assessment Problems

To gauge the relative effectiveness of the various treatments, I ended each interview by asking the visitor to answer two near-transfer problems. Both problems were presented as schematic pictures of novel arrangements of lamps, wall, and blocking object, and the visitor was asked to draw and explain the pattern of shadows that would result. The first, shown in Figure 5, involved two lamps of novel color and in novel locations with respect to the object. The second, shown in Figure 6, involved only one lamp. None of the subjects had seen these exact configurations during the interview, so neither problem could be solved by memory alone.

Choice of Visitors

The extended length, cognitive intensity, and nonsocial nature of the interviews made them especially challenging to administer. In the pilot studies, several visitors who had been somewhat reticent about doing the interview chose to leave before completing the final assessments or began answering the questions without much serious thought. To avoid this problem, I chose to compromise on random selection, by giving a general invitation to a group of visitors and letting them choose who would most like to do an interview. I also offered to interview two or three people at once, if they agreed to sit apart and not share ideas with each other. This served to reduce

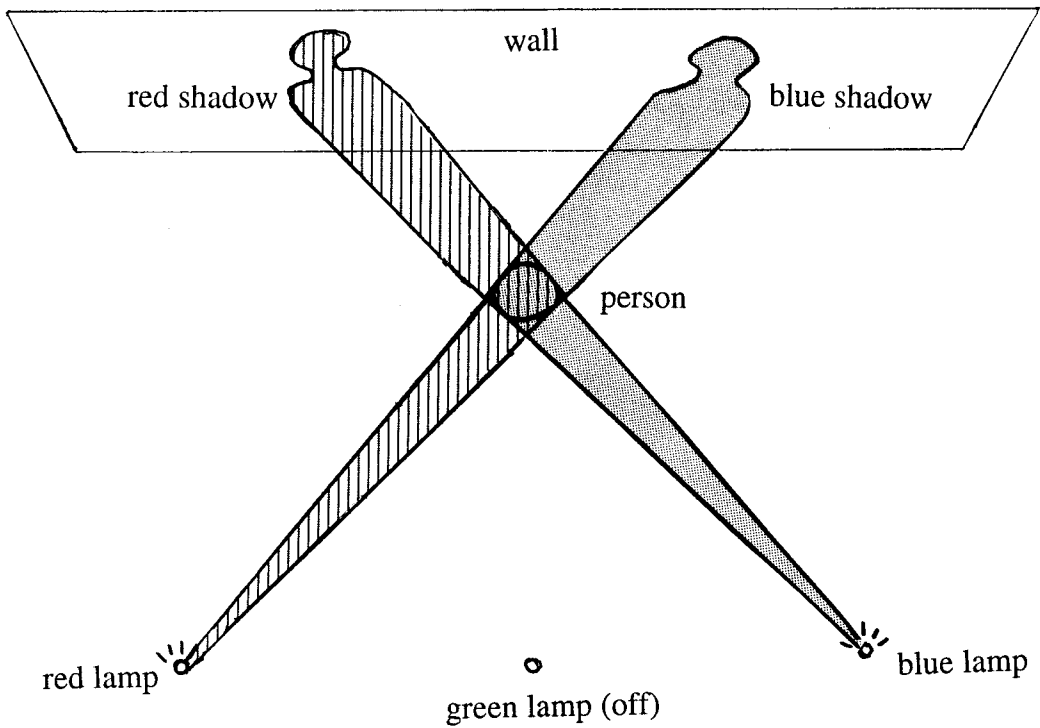


Figure 4. A common misconception (reflecting). Light from the lamps bounces off your body instead of going straight to the wall.

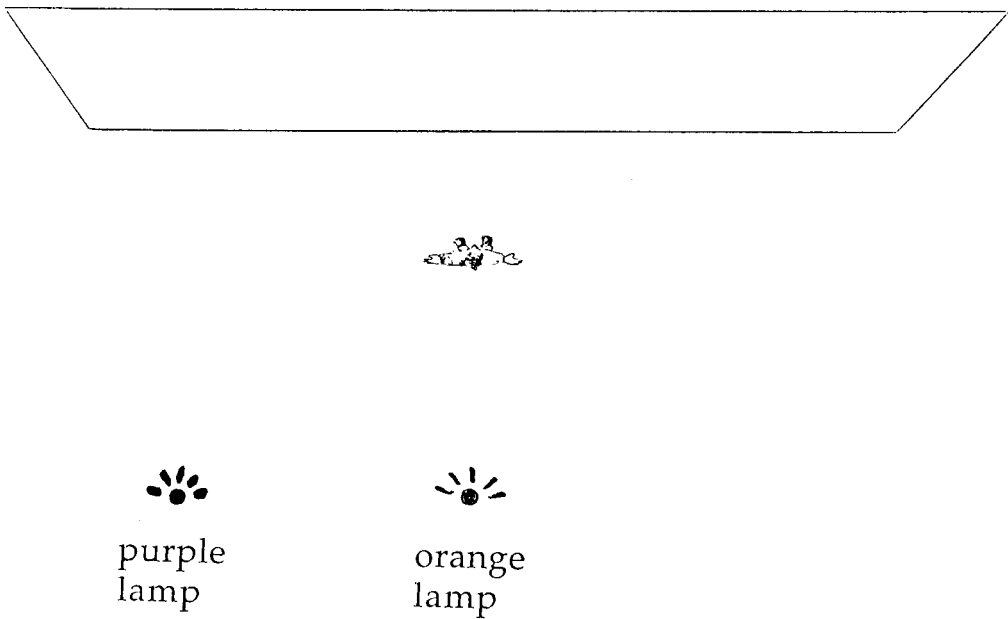


Figure 5. First transfer problem. Visitors were asked to draw and explain any shadows that would be seen on the wall.

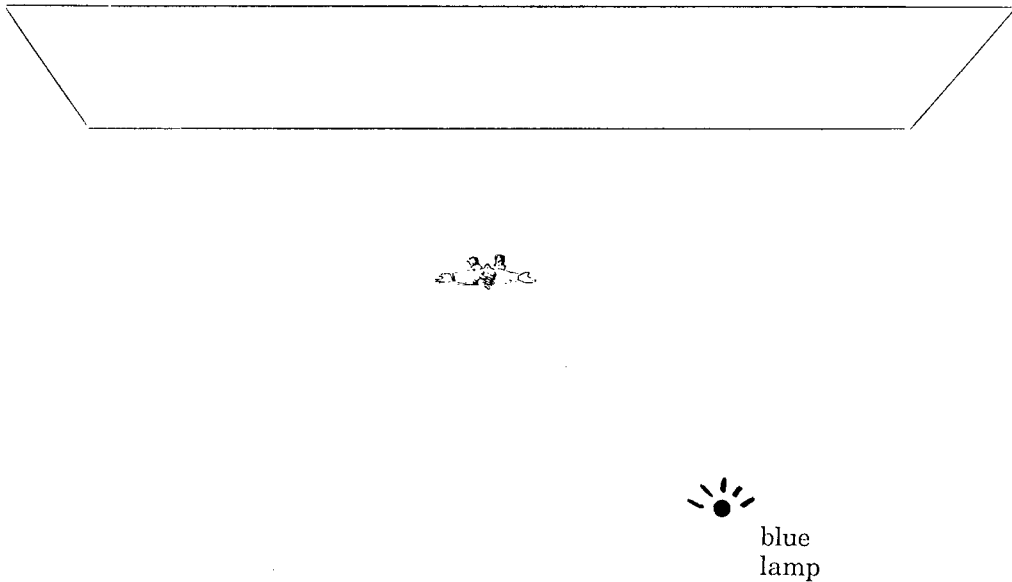


Figure 6. Second transfer problem. Visitors were asked to draw and explain any shadows that would be seen on the wall.

the frustration levels in groups where several people were eager to participate, whereas others preferred to watch or relax. To the extent that this may have introduced a selection effect, the reader may view the results of this study as a “best case” scenario, focused on visitors who have both the time and inclination to engage in inquiry activities in the presence of an interviewer.

In assigning individuals to specific inquiry activities, I used a randomized block design where each visitor drew a lettered ball from a box of balls *after* volunteering to do the interview, so that there was no possibility of bias toward any one of the inquiry conditions.

The final dataset contains equal numbers of males and females for each inquiry activity in each of five age categories. The total number of visitors interviewed is shown in Table 1.

RESULTS

Scoring

The two assessment problems were scored in combination, leading to a score of 1 or 0 for each visitor. A visitor scored 1 if and only if she correctly predicted the location and color of all shadows in both problems. The scoring criterion was thus quite strict; based on what visitors verbalized, I believe it is very unlikely that someone could answer both problems correctly without having a solid understanding of the physics behind the exhibit. On the other hand, this scoring criterion did not penalize a visitor who had good physical intuitions but poor articulation skills. Intercoder reliability tests showed agreement of 97% on the scoring.

Summary of Success Rate across all Inquiry Conditions

The data show the following characteristics, when summed over all inquiry conditions.

Age. Figure 7 shows the clear dependence of success rate on age. None of the 52 visitors younger than 13 was able to answer the transfer problems correctly. However, by age 16–18,

TABLE 1
Numbers of Visitors Interviewed

Age	Visitors Per Inquiry Condition	Inquiry Conditions	Visitors
7–9	4	7	28
10–12	4	7	28
13–15	4	7	28
16–18	4	7	28
Adult	40	7	280
Total for all ages	56	7	392

visitor success rates had reached adult levels of approximately 40% correct. There was no significant difference between success rates for the different age groups aged 16 and higher [$\chi^2(3) = 2.69, p = 0.44$]. For this reason, visitors aged 16–18 have been included with adults for the remainder of the data analysis.

Gender. Figure 8 shows the strong dependence of success rate on gender for nearly all age groups. Among visitors aged 16 and over, males were successful more than twice as frequently as females: 56% of males answered the problems correctly, as compared with 23% of females. This difference is significant [$\chi^2(1) = 50,000, p < 0.001$].

Evidence of Prior Understanding. As described earlier, I asked visitors early in the interview whether the arrangement of shadows seemed strange or sensible to them. Visitors who gave responses such as “It makes sense, because I’m blocking the red, so no red gets through,” or “If I think about it, you block the red, so you can’t possibly get a red shadow,” were coded as showing evidence of understanding prior to doing the inquiry activity. Visitors who said that the shadows seemed reversed from what they would expect, or who gave incor-

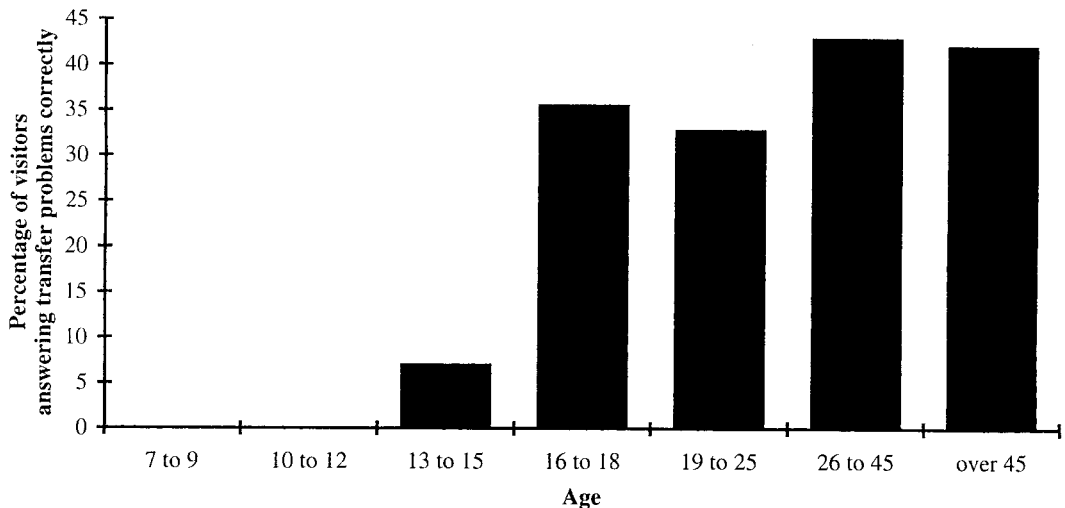


Figure 7. Success rate versus age ($n = 392$).

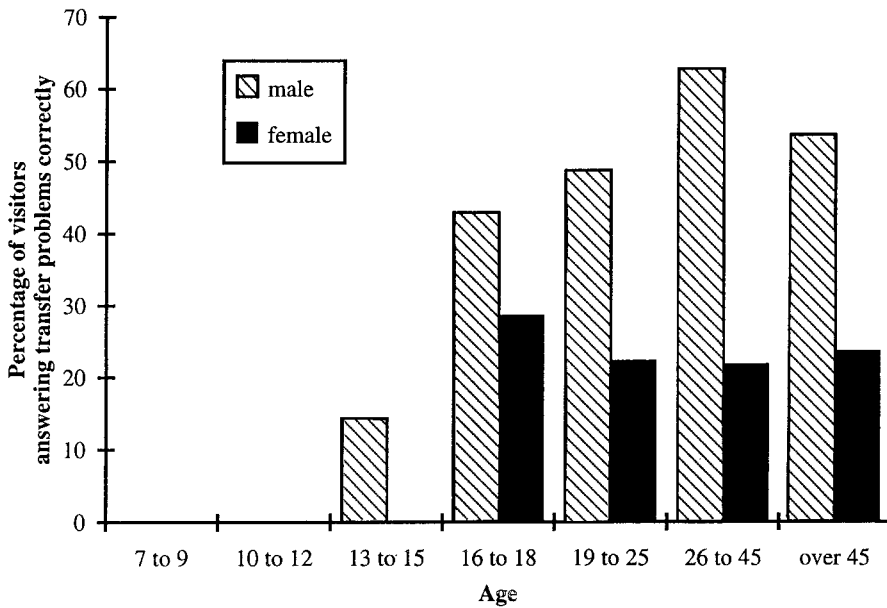


Figure 8. Success rate versus age for each gender ($n = 392$).

rect interpretations such as “It makes sense, because it’s reflecting off my back,” or “They aren’t crossing over; they stay on their sides,” were coded as showing no evidence of prior understanding. The only visitors who could not feasibly be asked this question were those in condition G, who made a prediction before seeing the phenomenon. For those visitors in conditions A–F, these self-reports proved to be a surprisingly good predictor of their final performance. Figure 9 shows the relationship between success rate and prior understanding as

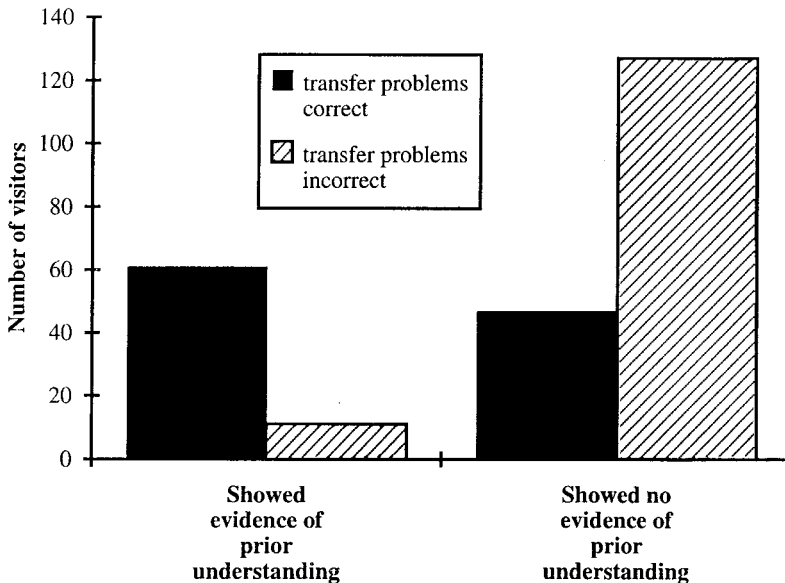


Figure 9. Dependence of success rate on prior understanding ($n = 246$ adults).

measured by the self-report question. Of visitors who, prior to the inquiry activity, showed evidence of understanding the role of light blocking in the exhibit, 85% answered the transfer problems correctly. Only 27% of those who showed no evidence of prior understanding were able to do so. These results show a dependence of success rate on prior understanding [$\chi^2(1) = 50,000, p < 0.001$].

Formal Education. Figure 10 shows that visitors' success rate was not significantly related to the amount of formal education, as measured by self-reported level of schooling [$\chi^2(3) = 2.27, p = 0.52$].

Formal Education in Science. As shown in Figure 11, success rate did correlate with the amount of visitors' formal education in science. Specifically, the transfer problems were answered correctly by 60% of those who reported having a degree in science, 43% of those who reported having had some college science, and only 25% of those who reported having had high school science or less. This pattern of results indicates an overall relationship between success rate and formal education in science [$\chi^2(2) = 18.49, p < 0.001$].

Effect of Inquiry Activity on Performance

Figure 12 shows the performance data for the 308 visitors aged 16 and over. The most successful visitors were those who did the standard explanation–interpretation activity (condition B). In this group, 57% of the visitors answered the transfer problems correctly, significantly higher than the percentages in groups A, C, and G, where visitors were asked to generate, troubleshoot, or predict first. In these latter groups, success rates were 34%, 34%, and 23%, respectively [$\chi^2(1) = 4.6, p = 0.03$; $\chi^2(1) = 4.6, p = 0.03$; $\chi^2(1) = 10.7, p = 0.01$]. Group B visitors were also more frequently successful than visitors who had done one of the three activities involving making a choice between two explanations (viz., D, E, F), although

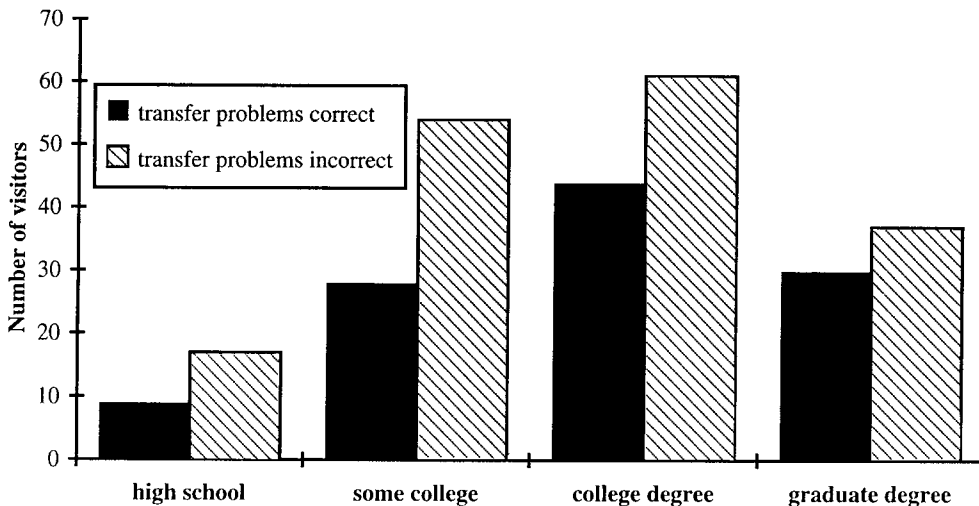


Figure 10. Dependence of success rate on formal education ($n = 280$ adults).

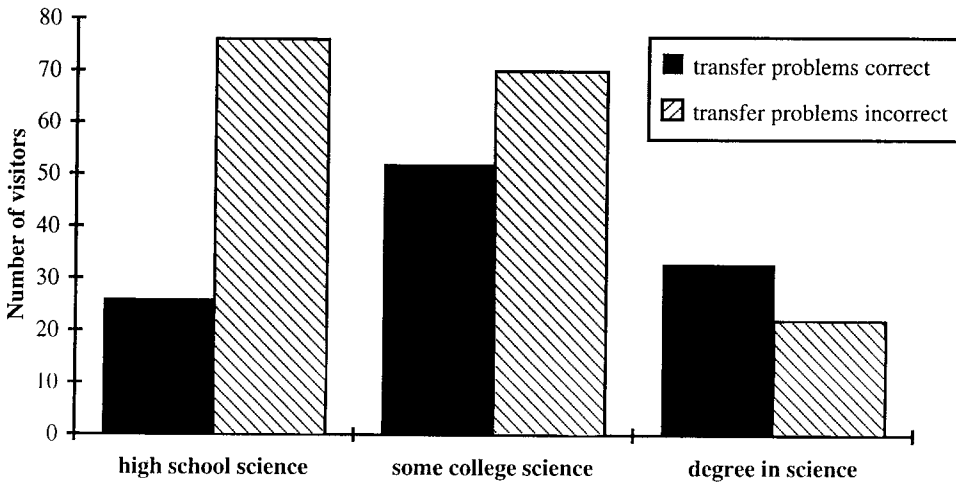


Figure 11. Dependence of success rate on formal science education ($n = 280$ adults).

these differences were not significant. In these groups, 45%, 39%, and 43% of the visitors answered the transfer problems correctly.

One might be particularly interested in performance differences among visitors who did not show evidence of understanding prior to the inquiry activity, because they are the ones displaying the greatest potential for learning. These data are shown in Figure 13. The same pattern of results describes this subgroup of visitors as the larger sample shown in Figure 12.

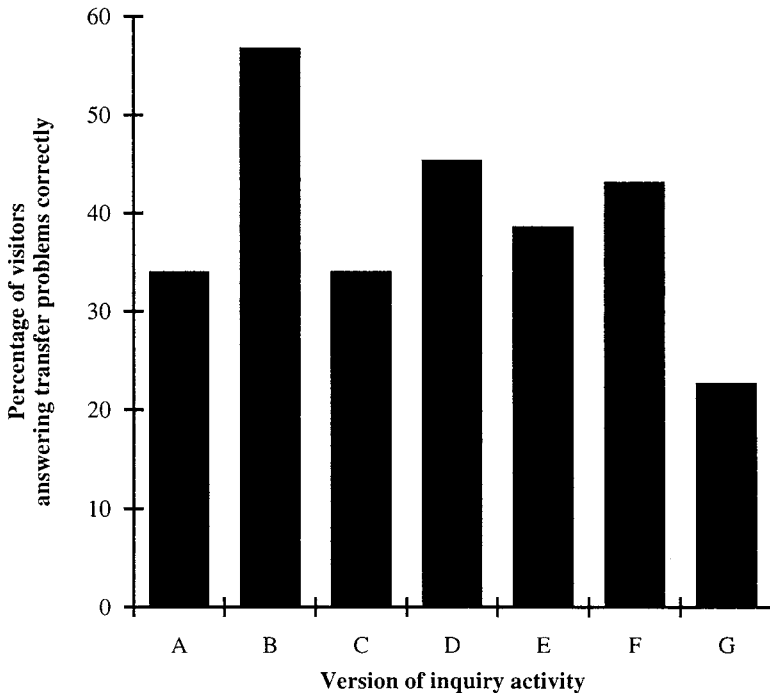


Figure 12. Dependence of success rate on inquiry activity for visitors aged 16 and over ($n = 308$).

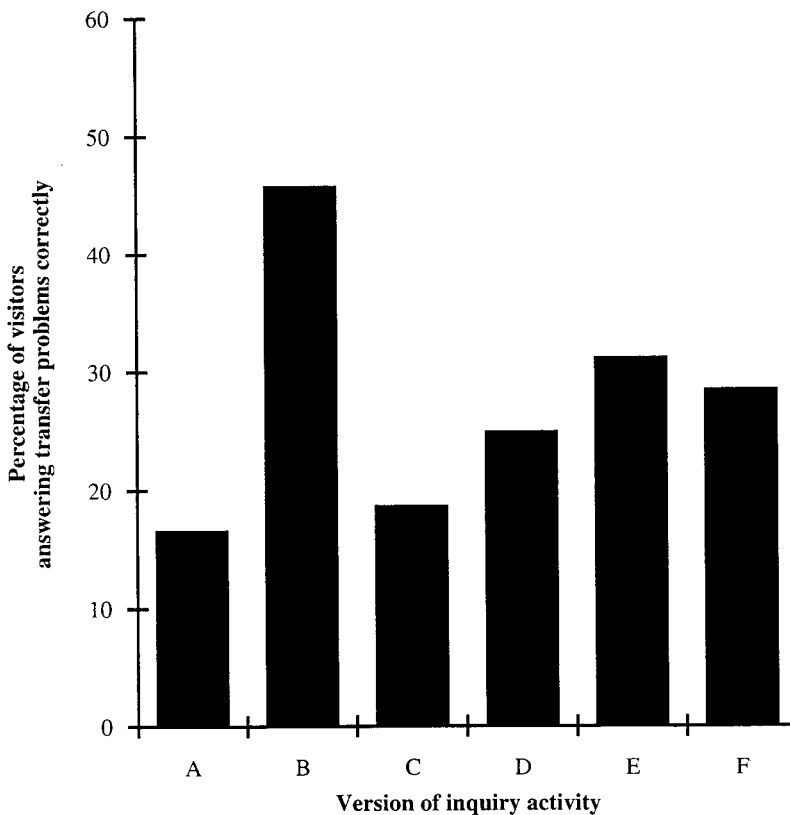


Figure 13. Dependence of success rate on inquiry activity for visitors aged 16 and over with no previous knowledge ($n = 174$).

Specifically, the percentage of visitors who answered the transfer questions correctly was significantly higher for group B than for groups A and C (with prior understanding being unmeasured for group G). The percentage correct was also higher for group B than for groups D, E, or F, but these differences were not statistically significant.

Range of Reasoning Shown in Visitors' Responses

Although visitors' responses to the transfer problems showed great variety, there were key commonalities. The majority of visitors answered the two problems by making some prediction about the paths of light from each lamp to the person and to the wall. In doing so, they often evoked ideas about blocking or reflecting, but many used variations of these ideas, and some even used hybrids that involved both at the same time. Figures 14–16 show examples of visitors reasoning with these different models while answering the first transfer problem. (In each case, the figure shows what the visitor drew, as well as what she said by way of explanation.) The hybrid reasoning shown in Figure 16 is particularly interesting. In this case, the visitor begins by reflecting the purple light to the left, a common response consistent with a reflecting model. However, the orange lamp presents a new challenge. Its central location makes reflection a difficult move to justify: would the orange light reflect to the left or right? This visitor decides, quite reasonably, to locate this second shadow in the center of the wall,

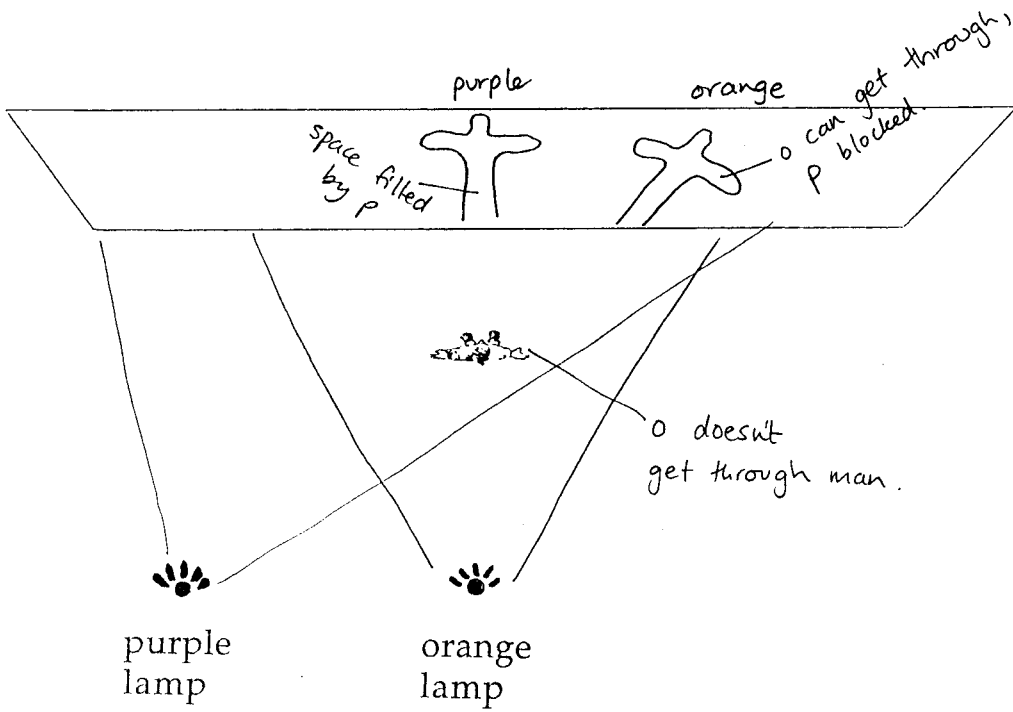


Figure 14. Example of visitor using "blocking" reasoning correctly.

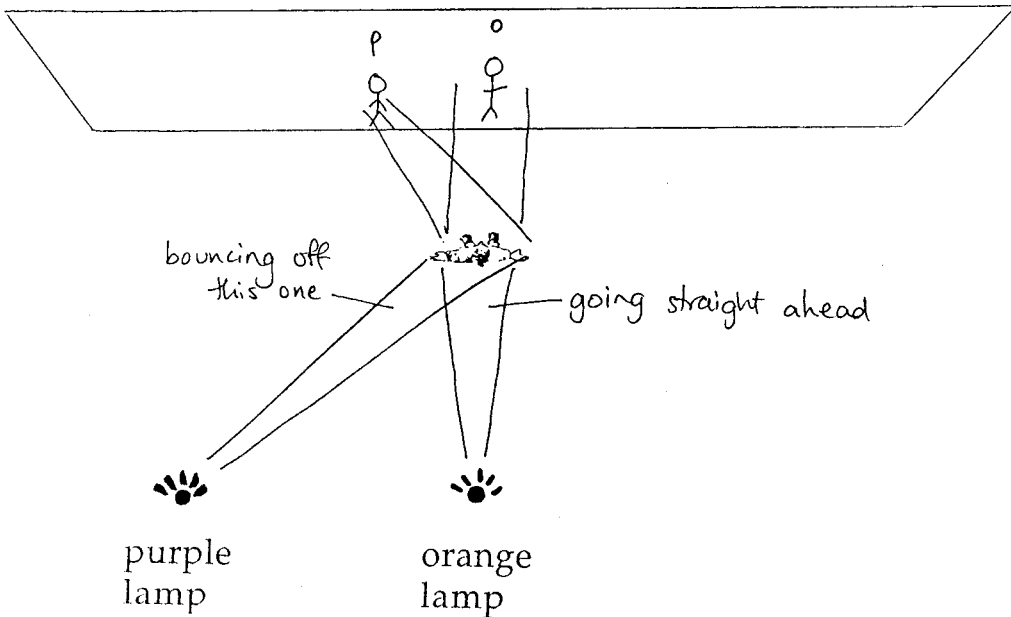


Figure 15. Example of visitor using "reflecting" reasoning.

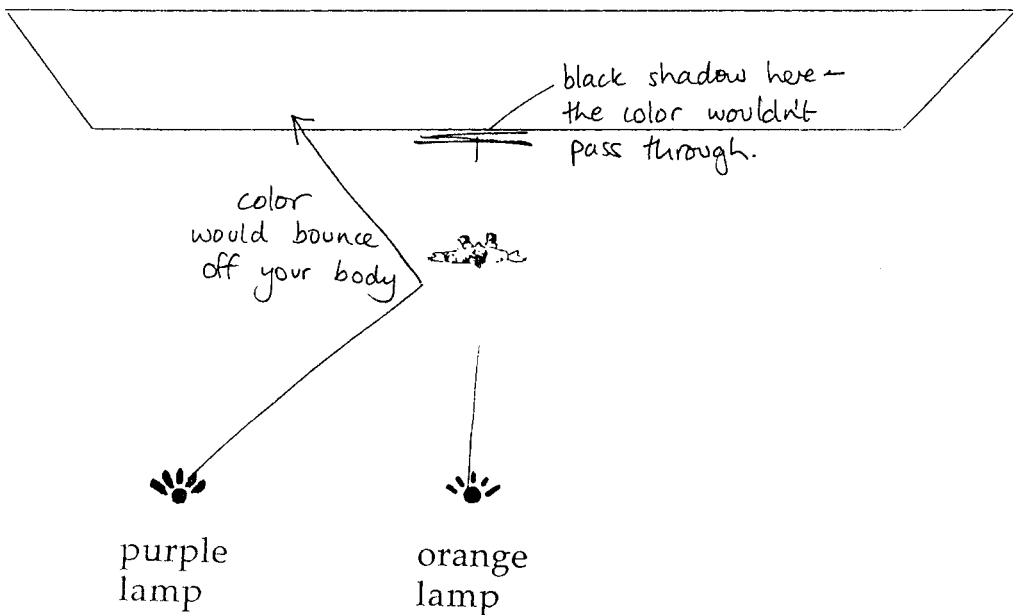


Figure 16. Example of visitor using hybrid reasoning: “blocking” with “reflecting.”

and at this point she invokes blocking reasoning, rather than allowing the orange light to go straight through the person. This example highlights the powerful nature of the first transfer problem to probe and challenge visitors’ understanding, through its asymmetry.

A minority of visitors, however, did not rely on modeling the paths of light. Instead, they reasoned by analogy, choosing certain features of the original situation that they felt would be preserved in the new situations. An example of this type of reasoning is shown in Figure 17. Finally, a small number of visitors reasoned through a process of transformation; they explained that the lamps would have to be moved in specific ways to get to the locations shown in the transfer problem, so the resulting shadows would be moved in similar ways compared with their current locations. This type of reasoning is exemplified in Figure 18.³

Interview Data

Up to this point, all results described have been based on visitors’ responses to the final transfer problems. However, some insights into visitor reasoning came from their reasoning processes during the earlier part of the interview.

(1) *Choosing the right explanation was relatively easy, but not necessarily powerful.* Four of the inquiry conditions (viz., D, E, F, G) involved visitors making a choice between one correct and one incorrect model of light behavior. Of the 176 people asked to make such a choice, 74% chose the correct explanation, based on blocking of light. However, the final performance of these 176 visitors averaged only 38%. In other words, most people were able to choose the correct model over a plausible competitor, but over half of them were left without a full understanding of that model. It seems that visitors were not constructing their own un-

³A more complete report of visitors’ diagrammatic reasoning and use of alternative models has been submitted for publication.

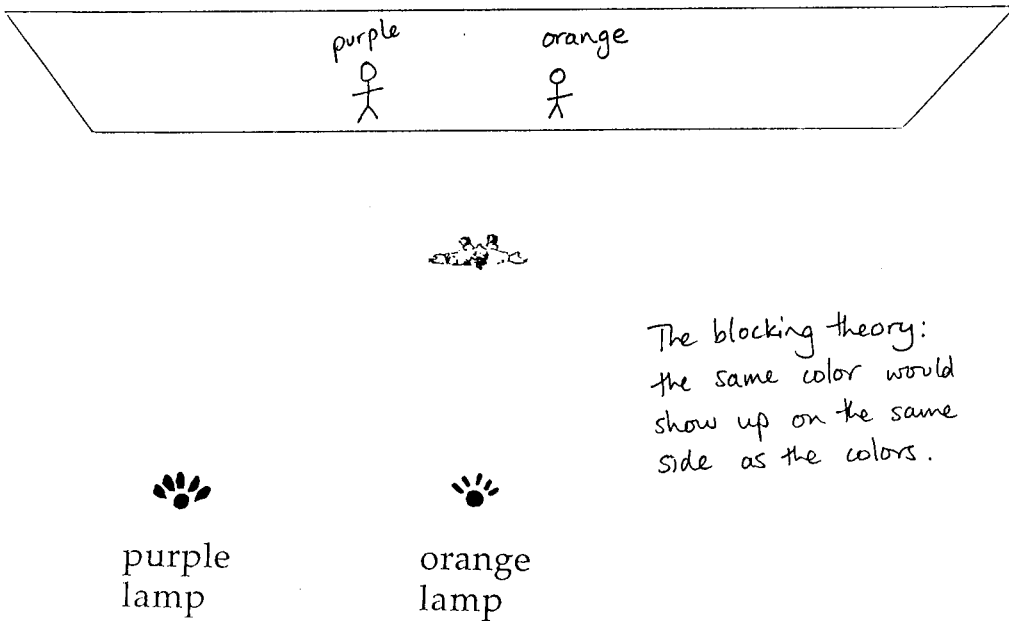


Figure 17. Example of visitor reasoning by analogy (contrary to her own labeling).

derstandings in quite the way we had intended. In fact, a substantial number (30% of those who did the choice-based inquiry activities, and 57% of those who did the prediction-first activity) answered the transfer problems using reasoning that was consistent with the model that they had just rejected, and inconsistent with the model they had just chosen.

(2) *Successful revisions of thinking were rare.* Of the 45 visitors who chose the incorrect

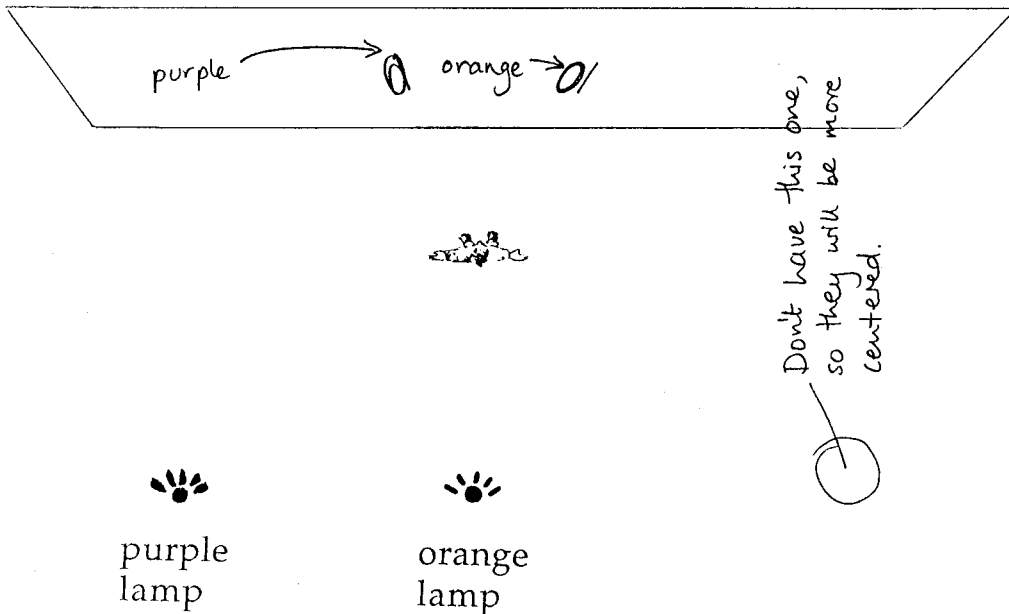


Figure 18. Example of visitor reasoning by transformation.

explanation over the correct one earlier in the interview, only 2 were able to solve the transfer problems correctly. This suggests that visitors with flawed understandings were not engaging in successful rethinking of their own ideas, even in the conditions which specifically encouraged such rethinking (conditions E, F, G). For example, in group G, 13 adults made predictions that were not borne out by experiment. In responding, 5 said that they had changed their minds and now believed the “blocking” explanation, but only one made a comment which substantiated this rethinking: “Oh, it’s the other way! So it blocks it from crossing over. It seems weird . . . I guess it intercepted the red, and so you blocked it.” The remaining 8 people either said they had no idea of how to interpret the surprising result, or turned to an incorrect model, such as reflection, to explain it. Of the total of 13 people who had experienced the opportunity and encouragement to rethink their understanding, only 1 answered the final two problems correctly, and more than half (8) used reasoning that was consistent with the reflection misconception.

(3) *Designing a discriminating experiment was relatively difficult.* Of the 44 visitors asked to design an experiment to discriminate between the “blocking” and “reflecting” explanations, 9 (20%) had no idea how to proceed. A typical comment was “No, I can’t think of anything. I bet there is a way, but I’m not a scientist.” A further 12 (27%) offered ideas that we could not try spontaneously within the existing exhibit. For example, visitors proposed moving the lamps, changing their colors, adding more walls, exploring with mirrors, darkening the room, and repeating the experiment with white lights. One visitor proposed viewing the shadow of a disk with a hole cut in it, but withdrew the suggestion when she realized it would not help to discriminate between the models. The remaining 22 (50%) did suggest some kind of experiment that could be tried within the exhibit constraints, and they were all invited to try out their proposals and interpret the results. Before each experiment, I asked the visitor to tell me more about how she thought it would discriminate between the two explanations. Interestingly, only 8 of the visitors were able to predict the outcome according to both models of light behavior (viz., reflecting and blocking). A further 6 could predict the outcome according to their own theory of choice, and 2 could predict the outcome according to the theory they hoped to disprove. Finally, 6 were not able to state any prediction at all, but suggested experiments as ways of exploring, in the hope that some insights would result.

There is some evidence that the experiment was more effective in facilitating learning if visitors had a clear idea of what they were expecting to see. Of the 9 people who showed no evidence of prior understanding, but who made specific predictions or could clearly see the discriminating properties of their experiment, 5 were later able to answer the transfer problems correctly. By contrast, of the 5 people who lacked previous knowledge, but proposed experiments without a clear prediction, none answered the transfer problems correctly. This difference is marginally significant under a bidirectional Fisher Exact Test ($p = 0.06$).

CONCLUSIONS

One of the things this study has shown is how much is involved in constructing a full understanding of the behavior of light in the colored shadows exhibit, even based on its simplified two-lamp form. Although the individual ideas underlying the exhibit are not highly technical, their coordination into a flexible predictive model is clearly a challenging task. Visitors younger than 13 seemed unable to construct such a mental model correctly, and even adults, concentrating throughout an individual interview, reached a success rate of only 39% (averaged across inquiry tasks).

Comparisons of these data with those from other studies are problematic, because of the high requirements for correctness in this study (viz., simultaneous correctness of color and lo-

cation on both problems), and the unusual difficulty of the near-transfer problems used. In particular, the first transfer problem (shown in Fig. 5) presents a configuration of lamps not previously reported. The combination of one lamp in the center, and one to the side, breaks the traditional symmetry of most transfer problems. As Figure 16 suggests, this asymmetry provided a challenging probe to visitors' understanding; it also revealed a variety of visitor diagrams and hybrid reasoning models.

Although many visitors constructed personal understandings of the exhibit that were not altogether correct, their thinking generally showed consistency and logic. The most common types of reasoning used were causal models that traced the path of light from each lamp to the object and to the wall. Such reasoning was most often based on ideas of reflection of light, blocking of light, crossing over of shadows, and reflection and blocking acting in combination.

The best predictor of a visitor's understanding, as measured by performance on a set of transfer questions, was his understanding of the role of blocking of light in creating the shadows, prior to beginning any kind of inquiry activity. Visitors with such prior understandings tended to be those who had taken college-level science courses. They were predominantly male.

The comparisons among inquiry tasks showed that the most effective inquiry activity was the explanation–interpretation task; the success rate of 57% for adults is quite impressive, given the difficulty of the transfer problems. The cluster of activities involving explanation–choice was less effective, although the differences were not significant at the $p < 0.05$ level. Significantly less effective were the inquiry activities A and C, in which visitors were asked to construct explanations, but were never shown the correct explanation in any form. It seems likely that visitors in these groups performed less well because, if they had constructed an explanation based on a plausible misconception (of which there are many), there was nothing to challenge their view. Another interesting result was the significantly poorer performance of visitors who were asked to make a prediction *before* seeing the intriguing phenomenon; this provides some support for the style of exhibit labels that start by leading the visitor through an experience, and explain it afterwards. Part of the problem with activity G (and also apparent in E and F) may have been that the role of concrete disconfirming evidence was much less powerful than anticipated as a motivator of deep revision of visitors' ideas. Presumably, the large cognitive effort of rethinking a set of linked ideas made this an unattractive option to many visitors. While this result is consistent with the extensive literature on the robustness of misconceptions in science, it does highlight the difficulties of facilitating conceptual change on the time-scale of a museum visit, and challenges us to explore further.

In a follow-up study, I will examine the effects of incorporating the more successful elements of these inquiry activities into stand-alone exhibit labels, with no interviewer present. This will help to reveal the effectiveness of the different tasks in the more complex and authentic context of the museum floor. In particular, the explanation–choice activities seem worthy of further study, because their emphasis on competing viewpoints may be especially effective in sparking conversations between visitors who think about light in different ways.

Joshua Gutwill, Mary Miller, and two anonymous reviewers provided very helpful comments on earlier versions of this paper.

REFERENCES

- Arons, A. B. (1990). *A guide to introductory physics teaching*. New York: John Wiley & Sons.
 Brown, A. L., & Campione, J. C. (1990). Interactive learning environments and the teaching of science

- and mathematics. In A. diSessa, M. Gardner, J. Greeno, F. Reif, A. Schoenfeld, & E. Stage (Eds.), *Towards a scientific practice of science education*. Hillsdale, NJ: Erlbaum.
- Burbules, N. C., & Linn, M. (1991). Science education and the philosophy of science: Congruence or contradiction? *International Journal of Science Education*, 13, 227–241.
- Feher, E. (1990). Interactive museum exhibits as tools for learning: Explorations with light. *International Journal of Science Education*, 12(1), 35–49.
- Feher, E., & Rice, K. (1986). Shadow shapes. *Science and Children*, 24, 6–9.
- Halloun, I., & Hestenes, D. (1987). Toward a modeling theory of physics instruction. *American Journal of Physics*, 55(5), 440–454.
- Hein, G. (1995). *Constructivism and museums*. Paper presented at the meeting of the Association of Science–Technology Centers, San Diego, CA.
- National Research Council. (1996). *National science education standards*. Washington, DC: National Academy Press.
- Perry, D. L. (1989). The creation and verification of a development model for the design of a museum exhibit (Doctoral dissertation, Indiana University, 1989). *Dissertation Abstracts International*, 50, 12A.
- Pfeiffer, B. (1995). *Interactive science exhibits on color concepts: Testing an educational design model*. Paper presented at the meeting of the National Association for Research in Science Teaching, San Francisco, CA.
- Pfundt, H., & Duit, R. (1991). *Bibliography: students' alternative frameworks and science education* (3rd ed.). Kiel, Germany: IPN.
- Science Curriculum Framework and Criteria Committee. (1990). *Science framework for California public schools: Kindergarten through grade twelve*. Sacramento, CA: California Department of Education.
- White, B. Y. (1993). ThinkerTools: Causal models, conceptual change, and science education. *Cognition and Instruction*, 10(1), 1–100.