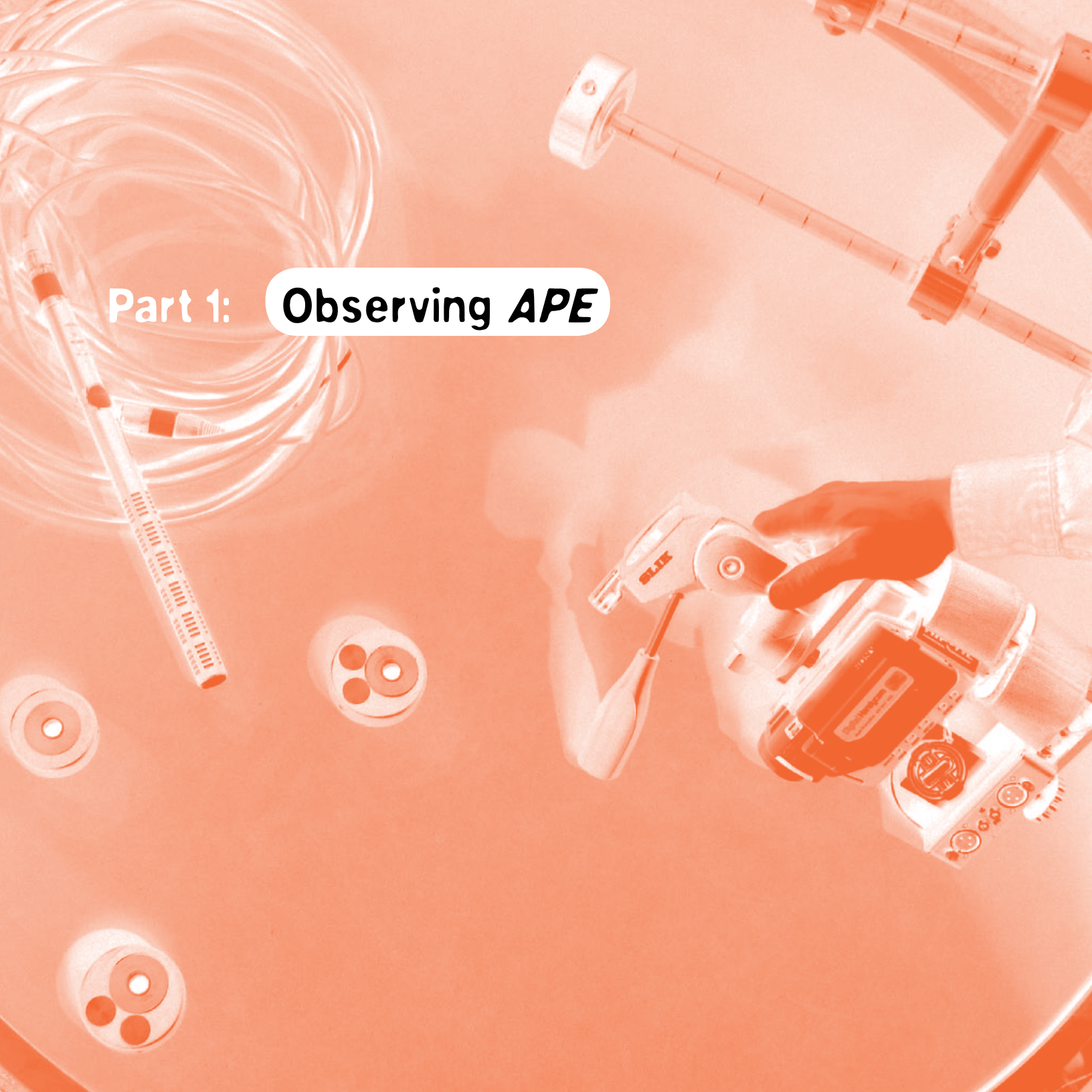


Part 1: Observing *APE*

Joshua Gutwill

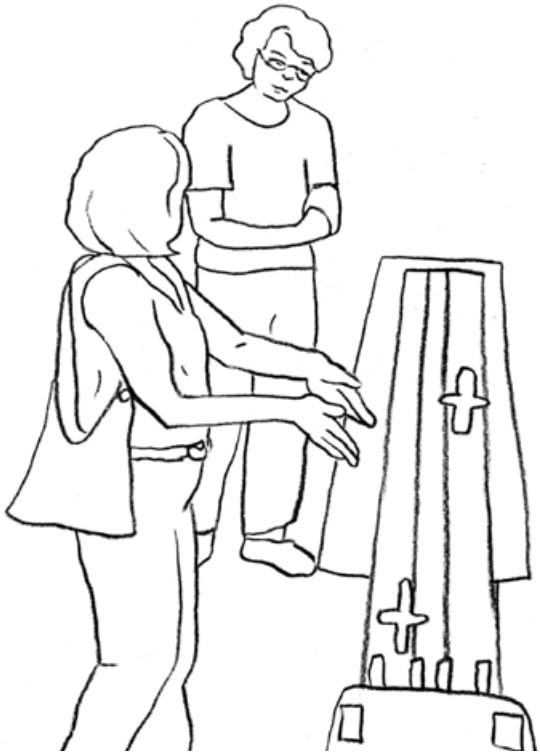
Converging on a Definition of *APE*

When we began the *APE* project, we weren't entirely sure what *active prolonged engagement* meant. We knew that we wanted *active* visitors to feel as if they were in the driver's seat, deciding for themselves what to try next, rather than following a set of instructions from the museum. By *prolonged,* we knew we wanted visitors to spend more time with the exhibits, getting more involved with the phenomena than they might at other exhibits. And finally, we had the idea that an *engaged* visitor group would be trying a variety of things at the exhibit, and that each action they took would somehow build on their previous actions.

To see and measure these components of *APE,* we realized that we would have to employ a set of evaluation methods that could capture visitors' thoughts and interactions. Audio/videotaping visitors while they used our exhibits seemed to be the best approach, because we would be able to see what visitors were doing and hear what they were saying. Still, we didn't know how to define *active prolonged engagement* well enough to measure it; in other words, we needed operational definitions of these ideas. We began by watching videotape of visitors at existing Exploratorium exhibits and early project prototypes. So, beginning by discussing moments in the videos that seemed potentially fruitful or interesting, we began to home in on a definition of *APE.*

One of the first steps in this process was to decide what kinds of interactions counted as *APE* behavior. Early in the project, the team disagreed about the kinds of actions that actually constituted *APE.* Did *APE* require some kind of focused investigation, inquiry, or hypothesis-testing by visitors, or could it be more broadly defined as free-form exploration of the range of possibilities offered by an exhibit? Did building complex structures from exhibit components count as *APE?* Were visitors' efforts to meet challenges or create aesthetically pleasing outcomes evidence of *APE?*

This debate arose after we began watching videotape of visitors at an early version of the *Downhill Race* exhibit. (See page 61 for a discussion of the development of this exhibit.) In group after group of visitors, we saw wonderful investigatory behavior: people making predictions, generating and testing hypotheses, and drawing conclusions. For example, two visitors, Shelley and Carol (all visitor names are pseudonyms), ran several "trials" to test their theory that air resistance might be a key factor in the race between disks rolling down an inclined path. In their first race, Carol compared a wheel containing holes with a thinner wheel without holes. She predicted that the perforated wheel would go faster—until she saw the thinner wheel pull ahead:



Visitors test their hypotheses at Downhill Race.

Carol: I would think that the, uh [wheel with holes would go faster].
Shelley: Yeah, the one with the holes will have less resistance, right?
Carol: Yeah.
Shelley: You would think so, but maybe this is [faster] because it is streamlined; see how thin it is, like cars, more streamlined. This [perforated wheel] is fat.
Shelley: OK, if you think that, OK, let's see if our theory is right. We will take another roller, one that's fat.

For their second race, they compared a wide wheel with holes and a wide solid wheel (thereby controlling for wheel thickness). Video footage like this—showing visitors making and testing hypotheses—excited several team members with the notion that we should build exhibits at which visitors could do genuine inquiry. However, others on the team felt that characterizing *APE* as primarily *investigative* behavior was too restrictive; they felt that visitors could demonstrate *APE* behavior by exploring a phenomenon in an engaged way, without making predictions, generating hypotheses, or drawing conclusions. One of the *APE* exhibit developers, Charles Sowers, summed up the issue:

There seem to be at least two approaches the visitor might take. One is analytical, where one follows a particular, reasonably well-formed line of thought in the attempt to arrive at a conclusion. This might be called investigative. The other is visceral or aesthetic, where one follows a chain of actions in the attempt to arrive at interesting or beautiful results or to test the limits of the phenomena. This might be called exploratory. The particular approach visitors will take and feel most comfortable with will depend largely on their own personal styles.

Like-minded team members presented videotape showing visitors engaged in exploration rather than investigation. Together, we watched a clip of a woman putting together colored tiles at the *Tiling Table* exhibit. Working alone, she said nothing during the experience, but she was clearly engaged. And there was no clear investigative purpose to her actions; she seemed simply to want to make a beautifully colorful, symmetrical object. At one point, she added a diamond-shaped tile onto her creation, then paused, rotated it, and moved it so that it was more symmetrically aligned with the other pieces. To many of us, this implied intentionality and suggested that she was not simply "doodling."

Watching this clip helped some members of the team accept *Exploration*, along with *Investigation,* as a form of active prolonged engagement. Eventually, two more forms of *APE* behavior emerged from watching video: *Observation* and *Construction. Observation* behavior surfaced at exhibits where visitors' main activity was carefully noticing the details of a phenomenon. Visitors seemed engaged for prolonged periods and seemed to be driving the interactions, but they weren't investigating or even widely exploring the phenomenon; instead, they were carefully and keenly observing. We became aware of

Construction behavior at exhibits where visitors built things like circuits, mathematical shapes, pulley systems, or mobiles. They often spent a long time developing their creations, seemed very involved, and drove the creating process themselves rather than relying on instructions from a graphic. We also noticed that visitors often left behind the artifacts they built for other visitors to see, augment, or dismantle.

The four types of *APE* interactions that emerged from watching video of visitors—*Exploration*, *Investigation*, *Observation,* and *Construction*—stretched our initial ideas about *APE* interactions, liberating exhibit developers to create exhibits in any of the four categories. They also helped us begin to clarify what we meant by *APE* as we noticed commonalities across the four. By building different kinds of prototype exhibits, watching and discussing visitor videos, and arguing over our ideas, we eventually arrived at the *attributes of APE* as a set of behaviors that we could look for when studying visitors.

Together, the four types of *APE* and the attributes of *APE* helped us think about the kinds of visitor interactions we would assess as we neared the end of the project. Rather than follow too closely the details of how visitors worked with any single exhibit, we would look for general signs of active prolonged engagement that could cut across several different exhibits.

Capturing Visitors' Interactions

Exactly how did we capture visitor interactions? How did we videotape visitors and then use the video to assess active prolonged engagement as we created exhibits?

Observing behavior at Tiling Table *reinforced* Exploration *as a form of* APE.

Attributes of *APE*

Who comes to the exhibit? How long do they stay? Why do they leave?

Visitors engaging in *APE* behavior:
- Show involvement with the exhibit despite coming from varied backgrounds;
- Spend more time and seem more involved with the exhibit than with other types of exhibits;
- Leave the exhibit for reasons unrelated or extrinsic to the exhibit (e.g., "we don't have much time") rather than intrinsic (e.g. "we did everything").

What questions do visitors ask? What do they talk about?

Visitors engaging in *APE* behavior:
- Ask their own questions, and use the exhibit to pursue answers, without relying fully on the authority of the museum.

How do visitors engage with the exhibits?

Visitors engaging in *APE* behavior:
- Talk to one another or give other indications that they are practicing scientific process skills such as inquiring, exploring, playing, observing, or contemplating;
- Continue interacting with phenomena even after reading the graphic's explanation;
- Try things suggested by exhibit graphics but not fully directed by them, or things entirely independent of graphics;
- Seem to be constructing a conceptual understanding of the exhibit phenomena.

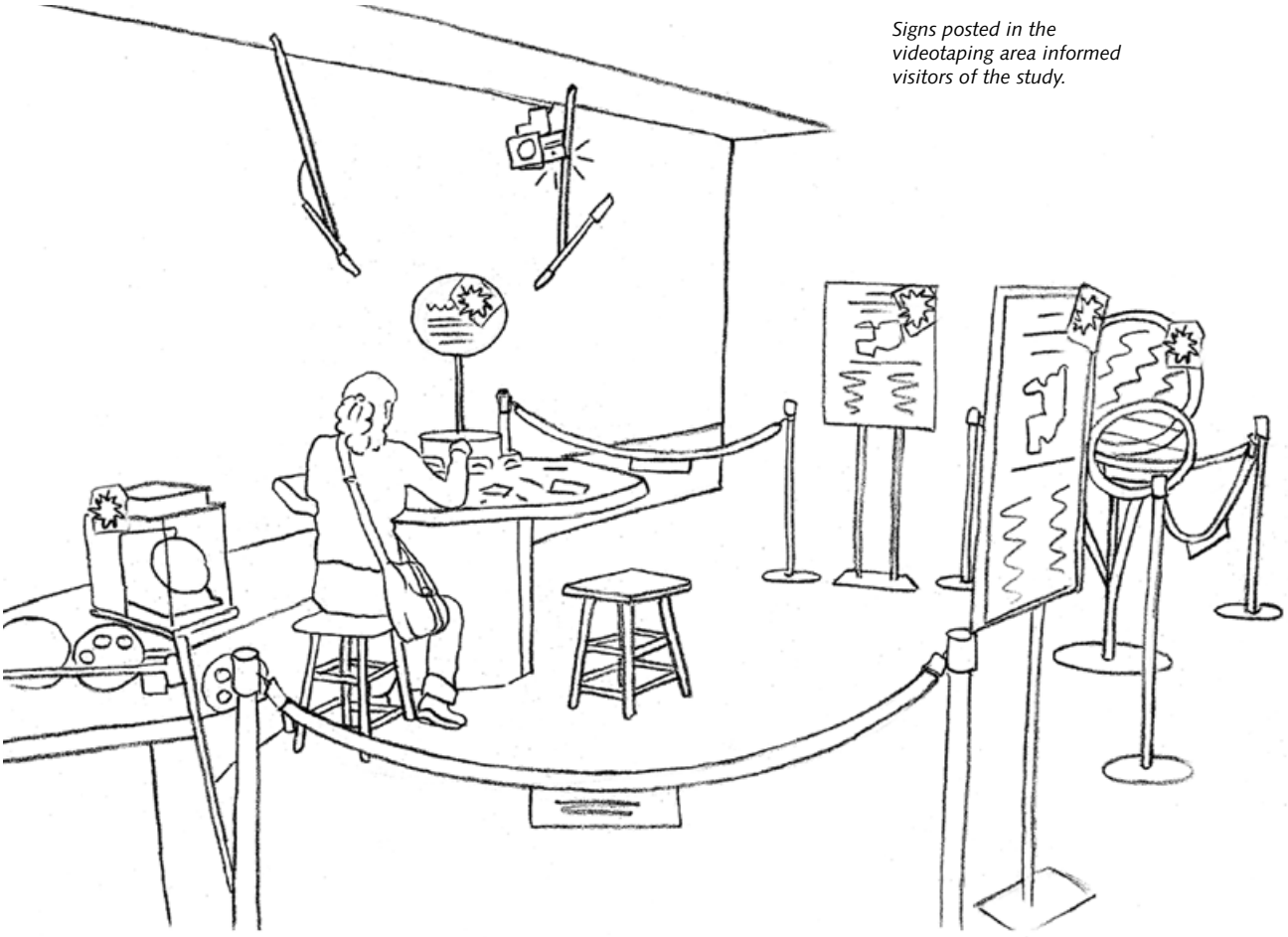
Preparing for videotaping

We began by making sure that we were protecting visitors' privacy, so we created a special videotaping area cordoned off from the rest of the museum floor. Following standard visitor-research practices, we posted signs at the entrances informing visitors of the research and of the videotaping. When we initially tested this setup, we found that only 75% of the 213 visitors we interviewed as they left the area knew they had been videotaped[1]. We then placed signs on all of the cordons and exhibits in the area, reminding visitors that they were being videotaped. A test of this new method found that 99% of the 200 visitors we interviewed knew they had been videotaped[2].

We also used sound-absorbing padding in the walls, ceiling, and floor of our videotaping area to reduce ambient noise. Along with our shotgun microphones, this allowed us to hear almost everything our visitors said, even in the cacophonous environment of the Exploratorium.

Selecting visitors to study

In most of our formative evaluation studies[3], we videotaped an exhibit for four hours on a single day, and then analyzed the first ten visitor groups from each hour, for a total of approximately forty groups per study. We often carried out such for-



Signs posted in the videotaping area informed visitors of the study.

mative studies iteratively, so some exhibits were taped in this way several times. In all, we collected over 300 hours of videotape at thirty-six exhibits and studied over 3800 visitor groups. (We used a subset of the videotape data for summative evaluation studies.)

Levels of detail in our video analyses

For formative evaluation, we analyzed the video data at one of three levels of detail, depending on the needs of the team. At the lowest level of detail, evaluators watched one hour of videotape and wrote a brief report noting the start and stop times for each group on the tape, describing visitors' behavior, and giving overall impressions of visitors' interactions. This quickly gave the exhibit developer a sense of the visitors' interactions with the exhibit, and helped the team determine whether the exhibit showed promise for further development.

At a medium level of analysis, evaluators found holding times for approximately forty visitor groups from across the four hours of videotape collected. Holding-time data were reported using graphs and descriptive statistics (mean, median, standard deviation, range, etc.). These data could also be compared statistically to previous versions of the exhibit, as well as to other Exploratorium exhibits. Museum researchers have debated the utility of holding time as an indicator of general visitor engagement, but we have found correlations between holding time and active prolonged engagement: groups who spent more time tended to show greater evidence for other attributes of *APE* than did groups who spent less time. Therefore, we used holding time as one indicator of *APE* behavior in visitors.

At the highest level of analysis, evaluators coded for specific visitor behaviors deemed important to the successful development of particular exhibits. For example, at *Spinning Patterns* (see page 29), we coded for both the number of patterns that visitors made in the sand and the nature of those patterns. We wanted to know whether visitors were intentionally making patterns or simply "doodling."

To do this, we created a set of categories for the kinds of patterns visitors might make—pre-pattern, simple pattern, or complex pattern—and then coded each visitor group for the most complicated pattern the group made. We also coded visitors' conversations to ascertain whether they discussed particular techniques for drawing patterns (such as "piling," "drawing," or "chopping"). Together, these codes gave the developer a sense of whether visitors were engaged in

Studying family groups

We chose the family and friends groups as the unit of analysis in most of our studies, rather than school groups or the individual visitor. We wanted to exclude school groups out of a concern that children's engagement would be driven by their peer group or by their teacher's assignment, skewing our results. According to visitor-studies research, people often visit museums in groups of family members or friends and they learn together as a unit. Choosing the family group as the unit of analysis meant that we calculated holding time from the moment the first person in the group approached the exhibit to the moment the last person in the group left. In our analyses of group interactions, we included and counted any actions made by anyone in the group, and we scored any overall physical and intellectual engagement for the entire group.

Visitors on good behavior

One concern we often heard from colleagues involved reactivity—visitors behaving differently because of the presence of camera and microphones. Certainly, we saw reactions to the camera; many visitors explicitly pointed out to one another the fact that they were being videotaped. In fact, we tried to make the camera, microphones, and posted consent signs as obvious as possible so that visitors would know that they were being recorded. Other museum researchers (such as Rennie and Johnston) rightly caution us to beware of visitor reactivity, but we felt that, because all of our studies involved videotaping, the reactivity would simply add a bit of noise to our signal. If visitors tried to look smarter for the camera, they would presumably behave that way at all the exhibits we studied.

carefully making patterns. (These codes were specific to the *Spinning Patterns* exhibit; other exhibits had other sets of codes for other interactions.) This highest level of analysis also included capturing group holding times.

Comparing Apples and Oranges

The goal of the project was to make exhibits that foster a different kind of visitor interaction from that engendered by other Exploratorium exhibits. We chose *Planned Discovery* (*PD*) exhibits as a foil for comparison. *PD* exhibits focus on specific ideas or concepts and often use a surprising phenomenon as a hook to get visitors interested in the idea: the museum has "planned" the exhibits so that visitors will "discover" the concepts.

When successful, *PD* exhibits are highly engaging and interesting to visitors. They evoke a positive emotional response, often evidenced by people calling to their family and friends to try the exhibit. Visitors' interactions with the exhibit are, however, usually limited to a small number of options; at many *PD* exhibits, there is only one option for visitors to try. Visitor behavior at these kinds of exhibits is often fairly predictable: visitors approach the exhibit, manipulate it, and are rewarded with a surprising outcome or effect. They might try it again or look more carefully at the phenomenon, but then they tend to turn to the exhibit graphic to learn about the underlying concept. After reading the graphic, they generally move on to the next exhibit.

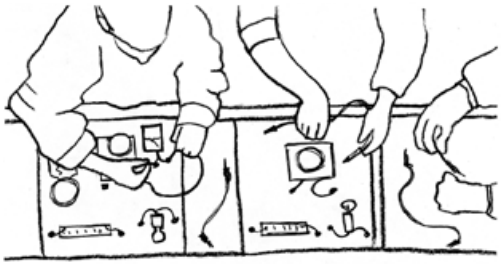
We envisioned that *APE* behavior would lie at one end of a continuum and *PD* behavior would lie at the other. Some exhibits in the *APE* project turned out to foster interactions that looked like the predictable *PD* cycle, and on closer study some *PD* exhibits fostered a few interactions that resembled *APE.* Generally, though, our summative evaluation studies were designed to determine whether our *APE* "apples" were indeed different from *PD* "oranges" in terms of how they affected museum visitors.

Assessing active engagement

Were the visitors more actively engaged at *APE* exhibits, asking and answering their own questions, or were they relying on the museum to instruct and inform them? Were they indeed acting more like participants than recipients in the learning process?

To answer these questions, we decided to compare the kinds of questions visitors asked at *APE* and *PD* exhibits, as well as the ways they answered those questions. We thought that if visitors were more actively engaged and curious, they would ask more questions while they used an exhibit. Active visitors might also ask more questions about an action to manipulate the exhibit (such as "What would happen if —?") and then use the exhibit to answer their questions. We thought that visitors who asked more *Why*—? questions would turn to the label for explanations, because it is often difficult to gain a satisfying explanation just by manipulating an exhibit. So the key evaluation questions here revolved around whether *APE* and *PD* exhibits evoked different kinds of questions for visitors and whether they encouraged different ways of answering those questions.

For example, consider a set of questions and responses among three young men at the *Circuit Workbench* exhibit as they try to understand a series circuit (see page 97 for more on this exhibit). They have just moved to a station that contains a generator, an incandescent light bulb, and a light-emitting diode (LED). After making a simple circuit between the generator and the LED, John asks whether they can make a circuit with all three components. They try connecting the generator, LED, and light bulb—but only the LED lights up.



Making connections, asking questions

John: But why does the light not work?
Steve: It needs more power.
John: Not enough power?
Ethan: Yeah, I think it needs more power.
John: What if you just go to the light instead of that [LED]?
Steve: If you go to the light it doesn't work.
John: What if you go through this one [incandescent light bulb] first?

They disconnect and reconnect everything, starting from the opposite terminal on the generator.

We analyzed the questions and responses of visitors like these at four *APE* and four *PD* exhibits. At each exhibit, we studied approximately 32 visitor groups, for a total of 252 groups. For each visitor group, we identified every utterance that sounded like a question (which we defined broadly as any request with a rising intonation at its end). We then categorized the question into one of five types of queries, as shown in Table 1. After we categorized visitors' questions, we categorized their responses to those questions into one of five groups as shown in Table 2[4].

We found that visitor groups asked three times as many questions at *APE* exhibits (8.3 on average) as they did at *PD* exhibits (2.7 on average). There was no difference in the number of questions asked per minute[5]. This means that visitors at *APE* exhibits spent more time at the exhibits and asked more questions than did visitors at *PD* exhibits.

We also discovered that visitors asked different kinds of questions at *APE* and *PD* exhibits. Visitors at the *APE* exhibits asked more Action and Explanation questions (48% at *APE* exhibits vs. 26% at *PD* exhibits), while visitors at the *PD* exhibits asked more Orientation, Perception, and Off-Task questions (52% at *APE* exhibits vs. 74% at *PD* exhibits)[6]. No matter what kind of question was asked, visitors at the

Table 1. Types of Visitor Questions

Question Type	Description	Examples
Action	Visitor asks for manipulation of exhibit elements.	Can you turn it faster? What if we connect this to this?
Explanation	Visitor asks for an explanation of exhibit phenomenon.	How does that work? Why did that happen?
Orientation	Visitor desires guidance about how to behave or use exhibit.	What's this one? What are you supposed to do?
Perception	Visitor asks for perceptions from others in the group.	Did you see that? Does it feel hot or cold?
Off-Task	Question is not about exhibit.	Did you know we're being taped? Are you hungry?

Table 2. Types of Visitor Responses

Response Type	Description	Examples
Reads Graphic	Visitor reads aloud, paraphrases, or silently reads graphic.	Q: Why do I see colors in the ice? R: These filters are polarized.
Use/Discuss	Visitor manipulates exhibit or discusses without referring to graphic.	Q: See how my hand is white? R: My nose is black.
No Response	Activity does not seem relevant to question asked.	Q: You know what I mean? R: [Other visitor doesn't respond.]
Off-Task	Almost always given when Question is categorized as Off-Task	Speaks directly into microphone. Manipulates microphone.

APE exhibits were more likely to answer the question by using the exhibit or talking with one another without referring to the label than were visitors at *PD* exhibits (79% at *APE* exhibits vs. 58% at *PD* exhibits)[7].

To give a sense of how these question and response categories differed between *APE* and *PD* exhibits, we show (see Table 3) how we would have coded the *Circuit Workbench* conversation transcribed on page 10. These young men never turn to the exhibit graphic to answer their questions, but try either building a circuit or explaining it to themselves.

For comparison (also in Table 3), we also present a coded conversation from a family group at a *PD* exhibit called *Touch the Spring,* where visitors reach into a box to touch a perfectly normal-looking spring, only to find that their hand goes right through it. (The spring is actually only an image of a spring, produced by a very smooth, large, curved mirror inside the box. There is also a flashlight at the exhibit, and the light makes the image look even more real.) At the start of our transcribed clip, the small girl in the group has just put the flashlight directly through the image of the spring, and her father asks her the first question. Notice how they turn to the exhibit graphic several times to answer their questions.

These results suggest that, at *APE* exhibits, visitors were more actively engaged in driving their own experience. They asked more questions; their questions focused on using or understanding the exhibit; and they relied less on the exhibit graphic to answer their questions than visitors at *PD* exhibits did. This last result was

Table 3. Comparison of *APE* and *PD* Behavior

Dialogue	**Question Type** (what visitors ask for)	**Response Type** (what visitors do)
APE* Exhibit: *Circuit Workbench		
But why does the light not work? It needs more power.	**Explanation**	**Use/discuss**
Not enough power? Yeah, I think it needs more power.	**Explanation**	**Use/discuss**
What if you just go to the light instead of that? If you go to the light, it doesn't work.	**Action**	**Use/discuss**
What if you go through this one first? *Disconnects and reconnects everything*	**Action**	**Use/discuss**
PD* Exhibit: *Touch the Spring		
Q: How did it go through there? R: Because this . . . *Mom points to graphic and reads aloud:* "When you shine the flashlight at the image of the spring, you're actually shining it directly at the curved mirror. The light beam bounces off the mirror and lights up the real spring inside the box."	**Explanation**	**Reads graphic**
It's like a trick, huh, Elsie?	**Explanation**	**No response**
Did you touch it? Touch it with your hand. *Girl reaches in with flashlight.*	**Action**	**Use/discuss**
It's not there. What? The spring.	**Perception**	**Use/discuss**
You know why it's not there? I'll show ya. *They move to the side of the exhibit and Mom uses a secondary graphic to explain:* Here's the real spring, and right here is where you put your hand through.	**Explanation**	**Reads graphic**

particularly encouraging: somehow, the *PD* exhibits were more likely than the *APE* exhibits to communicate to visitors that the exhibit graphic was of primary importance and should be consulted often. (Of course, reading graphics is not a bad thing per se, but it may reflect an attitude toward the museum as authority or instructor and therefore not encourage visitors to search for explanations on their own.)

In fact, we expected that *Why*—? questions would prompt visitors to turn to the exhibit graphics for explanations of phenomena, so we tried to encourage *What if*—? questions at *APE* exhibits so that visitors' attention would remain on the exhibits themselves. Surprisingly, even though visitors asked *more Why*—? questions at *APE* than at *PD* exhibits, they still tended to try to answer those questions by interacting with the exhibits rather than by reading the graphics.

Measuring "prolonged" holding times

We've shown evidence that visitors at *APE* exhibits were active, but were they engaged for prolonged periods of time? And what exactly does that mean? For us, it meant that visitors spent more time at *APE* than at *PD* exhibits. So we compared the average holding time for the fifteen exhibits described in this book with the average holding time for five successful and diverse *PD* exhibits and found that visitor groups spent three times longer at the *APE* exhibits: 3.3 minutes on average at the *APE* exhibits, compared with only 1.1 minutes on average at the *PD* exhibits[8]. (The average time spent at *PD* exhibits is in keeping with findings from other museums suggesting that visitors typically spend about one minute at an exhibition element.)

Of course, averages such as these tell only part of the story because most visitors spent very little time at either type of exhibit. However, at *APE* exhibits, 28% of visitors spent a long time—more than 4 minutes—while no visitor group we studied at a *PD* exhibit ever spent more than 4 minutes. (The longest recorded time spent at an *APE* exhibit was 58 minutes at *Circuit Workbench.*) Still, it's important to note that our goal was not to force everyone to spend a long time at exhibits, but to provide opportunities for prolonged engagement should visitors want it.

During the project, some team members began to raise doubts about our holding time measurement. Specifically, they became concerned that our prolonged holding times might be due less to *APEy* exhibits and more to *APEy visitors*—groups who would tend to spend time at *any* exhibit but were somehow more attracted to *APE* than to *PD* exhibits. Rather than appeal only to such visitors (should

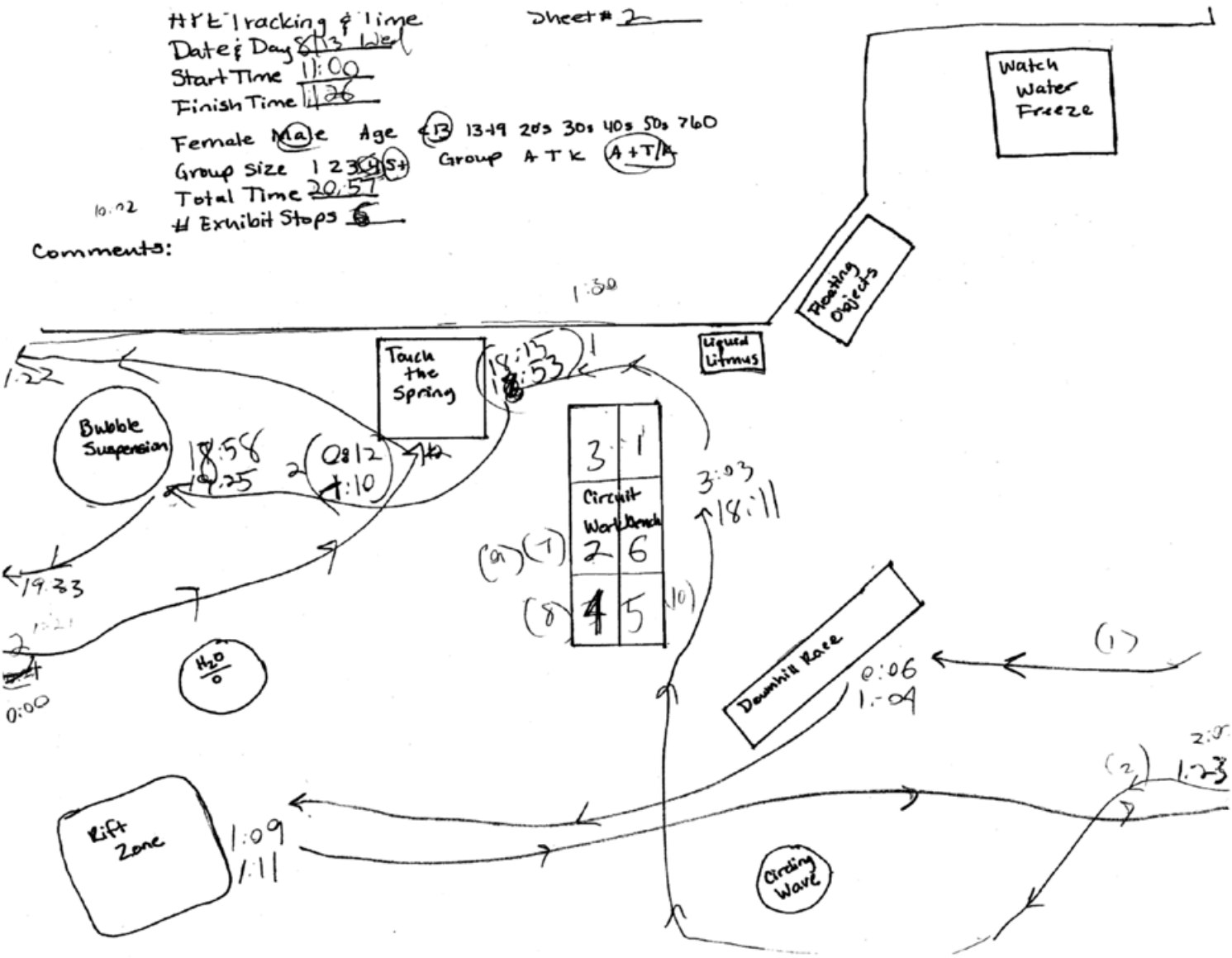
Sampling visitors in the question-and-response study

We sampled groups by taking the first eight eligible groups from each of the four hours of tape for each exhibit. An eligible group contained more than one visitor, and the group's members spoke English at least part of the time they were using the exhibit. By sampling the visitors who appeared early on each tape, we increased the likelihood that their entire visit would be captured on the sixty-minute videotape, even if they spent prolonged periods of time at the exhibit.

Controlling variables

One of the problems we faced in conducting comparative studies was that of controlling variables, such as location within the museum, context of surrounding exhibits, and time of day, month, or year. We controlled for placement and context by conducting all studies in our noise-reduction area, which was located approximately one-quarter of the way into the museum. Still, in the early part of their outing visitors were almost certainly acting "diligently" (exploring most or all exhibits rather than "cruising" and stopping at only scattered exhibits)—but, because that is true of all visitors in our studies, any differences we saw across exhibits should still be valid. To control for the context of the surrounding exhibits, we moved every exhibit from its contextual home and placed it in the noise-reduction area with two or three unrelated exhibits. We controlled for time by comparing different versions of the same exhibit on the same day. (This was not always possible, however, and we often had to resort to comparing different exhibits or iterations of the same exhibit across months or even years.)

Tracking and timing data sheet



they indeed exist), our team hoped to foster active prolonged engagement in *any* visitor to these exhibits.

To explore this issue, we employed tracking and timing techniques in which we observed 83 individual visitors as they moved through an area containing four *APE* and six *PD* exhibits. We timed each randomly sampled visitor at each exhibit and noted the paths each one took through the area. (A sample data sheet is shown on the opposite page.)

To compare the time that each visitor spent at *APE* exhibits with the time that the same visitor spent at *PD* exhibits, we calculated the average *APE* time and the average *PD* time for each visitor. The average *APE* time was calculated by adding up the times the visitor spent at all *APE* exhibits used and dividing by the number of *APE* exhibits used. The average *PD* time was calculated in a similar fashion.

We included only the 61 visitors who used at least one *APE* and at least one *PD* exhibit. Then, for all 61 visitors, we compared the average *APE* times with their average *PD* times and found that visitors spent more than twice as much time at *APE* exhibits (2.2 minutes) as at *PD* exhibits (0.9 minutes)[9]. This suggests that the type of exhibit, not the type of visitor, is likely to be responsible for holding time differences between *APE* and *PD* exhibits. Taking all these results into account, we concluded that *APE* exhibits were indeed more successful in promoting prolonged engagement than were *PD* exhibits.

Tracking individuals instead of groups

The tracking and timing study focused on individuals as the unit of analysis, because it is more difficult to track a group: when tracking a group, one must first determine who is in the group. This is possible when watching videotape, but it is nearly impossible to do in real time at busy exhibits. Besides, the purpose of our tracking and timing study was to compare *APE* and *PD* exhibits. Holding times would be measured for individuals at both types of exhibits, thus providing a fair comparison.

A family investigates delicate ice patterns at Watch Water Freeze.

Different kinds of engagement

The question-and-response and holding time studies told us that visitors seemed to be actively involved with *APE* exhibits for prolonged periods when compared to visitors at *PD* exhibits—but were they engaged in different ways? To answer this question, our external evaluators, Deborah Perry and Carey Tisdal of Selinda Research Associates, conducted a summative evaluation of visitor behavior at *APE* exhibits. During her doctoral work and as part of her summative evaluation of the *Experiment Benches* project at the Science Museum of Minnesota, Deborah Perry had already developed a framework for assessing in-depth engagement at museum exhibits. She modified that framework for the *APE* project.

Perry and Tisdal used an in-depth, naturalistic approach in which they observed

and interviewed visitor groups at three *APE* and three *PD* exhibits and then described each group's physical, intellectual, social, and emotional engagement. In this fashion, they hoped to capture the range of ways visitors engaged with *APE* or *PD* exhibits. They found that visitor engagement differed at *APE* and *PD* exhibits[10].

For example, physical engagement seemed quite uniform across *PD* exhibits, where visitors would engage in a predictable "do, notice, read" cycle. But at *APE* exhibits, physical engagement was more prolonged and varied from one exhibit to the next. Intellectual engagement at *PD* exhibits focused on knowledge about the phenomenon being displayed, while visitors at *APE* exhibits tended to think not only of the phenomenon but also of the process of exploring it. Social engagement differed as well: at *PD* exhibits, dyads tended to split off from bigger visiting groups to use the exhibits, while larger visitor groups were often able to use *APE* exhibits together.

In measures of emotional engagement, visitors at *APE* and *PD* exhibits showed marked differences. At *PD* exhibits, visitors consistently felt "positive, pleasant, and playful" and enjoyed the "trick" or "surprise" aspect of the exhibit, but these positive feelings did not necessarily lead to further engagement. In general, these visitors wanted to know why the phenomenon had occurred, but did not feel stupid if they did not understand it. At *APE* exhibits, on the other hand, visitors seemed to display a wider spectrum of emotional responses, ranging from pride at building a complex structure, to pleasure in creating an attractive aesthetic experience, to frustration at being unable to master a difficult challenge. This greater breadth of responses suggests to us that visitors to *APE* exhibits were bringing more of themselves—their fascinations, desires, questions, goals, expectations—to the exhibits than were visitors to *PD* exhibits.

In general, it seemed that visitors were indeed engaged differently at the two kinds of exhibits. (In a follow-up study, Perry and Tisdal examined six other *APE* exhibits to clarify the types of visitor interactions they engendered[11].) From all our studies comparing *APE* and *PD* exhibits, it would seem that we did manage to create apples that were different from the oranges that already existed in the Exploratorium: our *APE* exhibit apples seemed better at fostering active prolonged engagement by visitors.

Why Did We Find Differences Between *APE* and *PD* Exhibits?

Using videotape analysis and in-depth interviews, we and Selinda Research Associates found that visitors seemed more actively engaged for prolonged periods at *APE* than at *PD* exhibits. But why did these exhibits have such effects? A definitive answer to such a question would constitute a formula for building successful exhibits to foster *APE* behavior—and we do not believe in foolproof prescriptions for exhibit development; too many details vary from one exhibit to the next. We do hope that the rest of this book will give enough of an answer to this question for the reader to begin building, modifying, and iterating new exhibits to promote *APE*. However, we did conduct a few studies that shed light on the design features of our exhibits which seem to promote *APE*.

APE exhibits seem more open-ended

The first such study was spawned by a developer's complaint that visitors seemed to get engaged with his exhibit but were then pulled away by someone else in the group. This irked him because "the exhibit seems engaging enough, but people are getting dragged away by their friends; they're not staying as long as they might. So how do we know whether the exhibit is actually working to promote *APE?*" We realized that there was a positive interpretation of this: visitors were leaving because of external factors rather than because there was some kind of endpoint built into the exhibit.

We decided to interview visitors as they left twelve of the *APE* exhibits featured in this book (and four additional *PD* exhibits) to find out what had prompted them to move on to the next exhibit. We collected about ten interviews per exhibit and

Table 4. Visitors' Reasons for Leaving Exhibits

Reason for leaving	***APE*** (*N* = 270)	***PD*** (*N* = 40)
Extrinsic to the exhibit	60%	40%
Intrinsic to the exhibit	18%	45%
Both	22%	15%

categorized visitors' responses into *extrinsic* factors (not related to the exhibit) or *intrinsic* (related to the exhibit). Here are some examples of each:

Extrinsic factors

"The museum is so big—so, time, really. So far, I think this one is the most interesting."

"Because my husband called me. It was time to move on. You cannot spend as much time as you want at each one."

"Someone else wanted to try it."

Intrinsic factors

"I felt like I had done everything that the experiment had to offer."

"I finally got it. I got what they were asking for. Well, OK."

"I saw the objective of the activity."

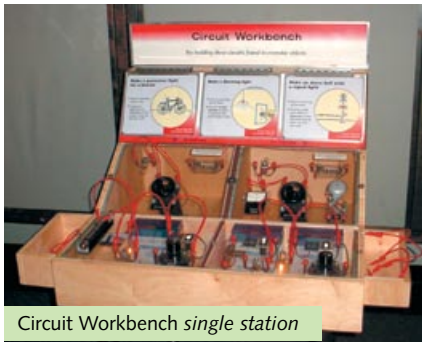
We found that more visitors to *APE* exhibits cited reasons for leaving that were extrinsic to the exhibit and more visitors to *PD* exhibits cited intrinsic reasons for leaving[12]. The percentages of visitors in both groups giving extrinsic, intrinsic, or both kinds of reasons for moving on to new exhibits are listed in Table 4.

The main issue underlying visitors' reasons for leaving, we believe, is one of open-endedness: presumably, visitors will tend to leave an open-ended exhibit for reasons extrinsic to the exhibit itself and will leave a closed-ended exhibit for reasons intrinsic to the exhibit. We designed *APE* exhibits to be open-ended, to have many options and few obvious stopping places. We hoped that an *APE* exhibit would never communicate to visitors that they were done, but would provide as much opportunity for engagement as the visitor desired. If engagement means being able to try different things, where one action follows from the next, engagement seems likely to depend on such open-endedness. We believe that visitors sensed the myriad possibilities at *APE* exhibits and that this affected their engagement.

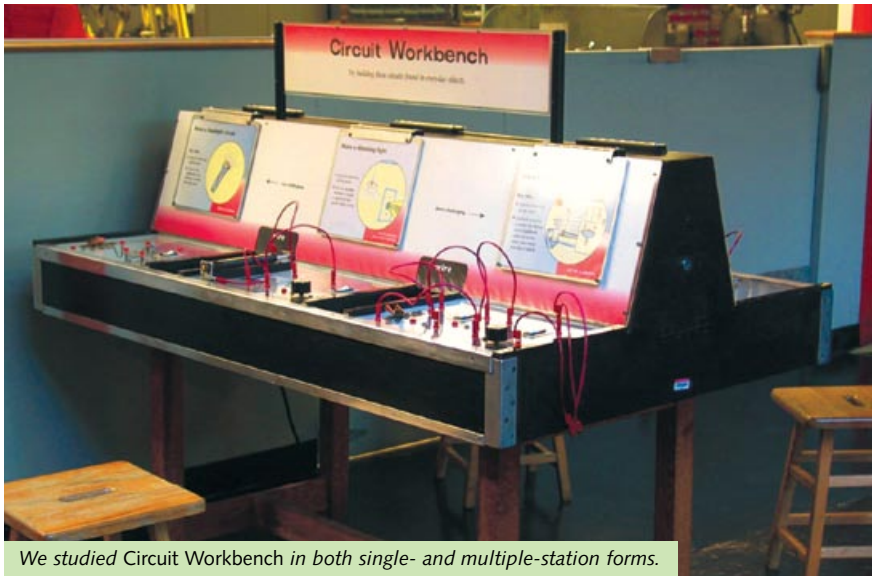
It is also interesting to note that some of the intrinsic reasons given for leaving seem to imply that the museum is an authority, telling visitors what to do (such as "I got what they were asking for"). Such responses subtly place the visitor in the role of information recipient and the museum in the role of instructor. So perhaps the open-endedness of *APE* exhibits also contributed to the active role that visitors took.

Studying multiple-station exhibits

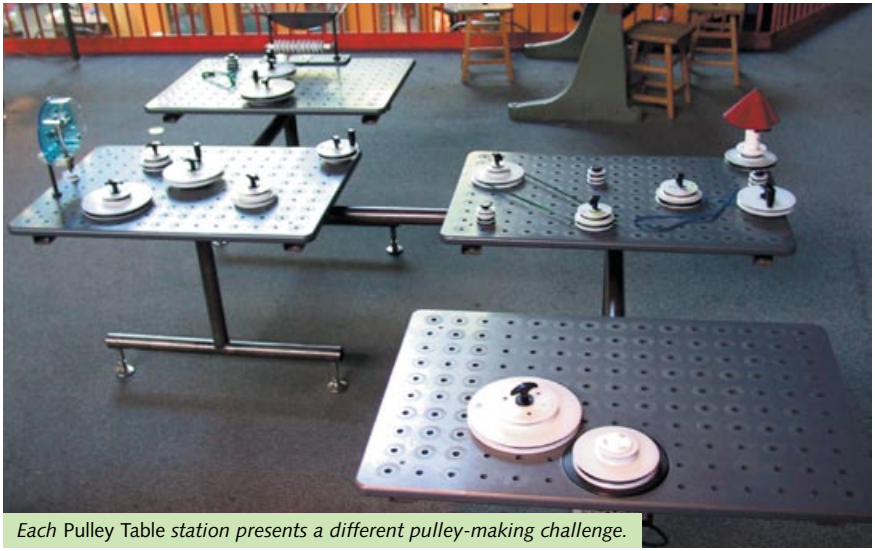
As our project and evaluation studies progressed, we noticed that some designs seemed particularly successful at fostering *APE*. One that seemed to work well was the multiple-station design in which an exhibit was divided into several interaction stations. At each station, one or two visitors could explore the phenomenon with total control of the components, but still use the exhibit as part of a larger group. Thus, social interactions seemed to be supported while the control needed for good engagement was also provided.



Circuit Workbench *single station*



We studied Circuit Workbench *in both single- and multiple-station forms.*



Each Pulley Table *station presents a different pulley-making challenge.*

Keeping a visiting group together also reduced the chances that one member of the group will call other members away from the exhibit. Minda Borun and others (1996, 1998) found that exhibits that promote what they called "family learning" are multisided and multihanded; in other words, they are *multiuser*. Building on Borun's work, we began to wonder if a key component of a multiuser exhibit is that each person using it has control over his or her own play space.

Single-station version of Pulley Table

We decided to see whether such a design really did change interactions and promote active prolonged engagement. We chose three *APE* exhibits and built them out in both single- and multiple-station formats, then compared visitor behavior at the different formats. Like many Exploratorium exhibits, the single-station version allowed several visitors to use the exhibit at once, but would also require them to share control of the exhibit.

We chose to work with *Spinning Patterns, Circuit Workbench,* and *Pulley Table* (all described in the *APE Tales* section of this book) because their multiple-station configurations differed from one another in an interesting way: At *Spinning Patterns,* the three stations were identical—each had a spinning platform, sand, a shovel, and a stick for drawing and erasing patterns in the sand. We figured there would be little incentive for visitors to get up and move from one station to another as each station provided the same tools and phenomenon. *Circuit Workbench,* in contrast, consisted of six stations, each offering different combinations of electrical devices. For instance, one station had a battery, a motor, a light, and a variable resistor; another had a generator, a light, LEDs, and a bell. Given these kinds of differences between stations, we thought that visitors to *Circuit Workbench* might move from one station to the next, thereby prolonging their engagement. At *Pulley Table,* visitors had similar tools at all four stations (pulley wheels and belts), but there were different implied goals at each of the four stations (use a motor, spin an umbrella, crank a fan assembly, or activate a music box). Here, we thought visitors might move among the different stations and even create pulley systems using several stations at once.

The phenomena and affordances for interacting with the exhibit remained the same across the single- and multiple-station versions. (We also compared behavior *across* exhibits, but this was a weaker comparison because so many other variables came into play, such as the number of stations or differences in the intrinsic appeal of each exhibit's core activity.)

At each version of each exhibit, we videotaped and interviewed fifty visitor groups (300 groups in all), focusing on what they thought about while using the exhibit and on why they left the exhibit. Again, the videotape analysis focused on the visitors' physical, intellectual, and social engagement. We also measured visitors' holding times. We found a marked difference between single- and multiple-

station designs only at the *Spinning Patterns* exhibit; we found few differences at either *Circuit Workbench* or *Pulley Table.* (Across exhibits, holding time correlated with our measures of physical, intellectual, and social engagement.)

At *Spinning Patterns,* visitors were more engaged at the multiple-station version by almost every measure we used:

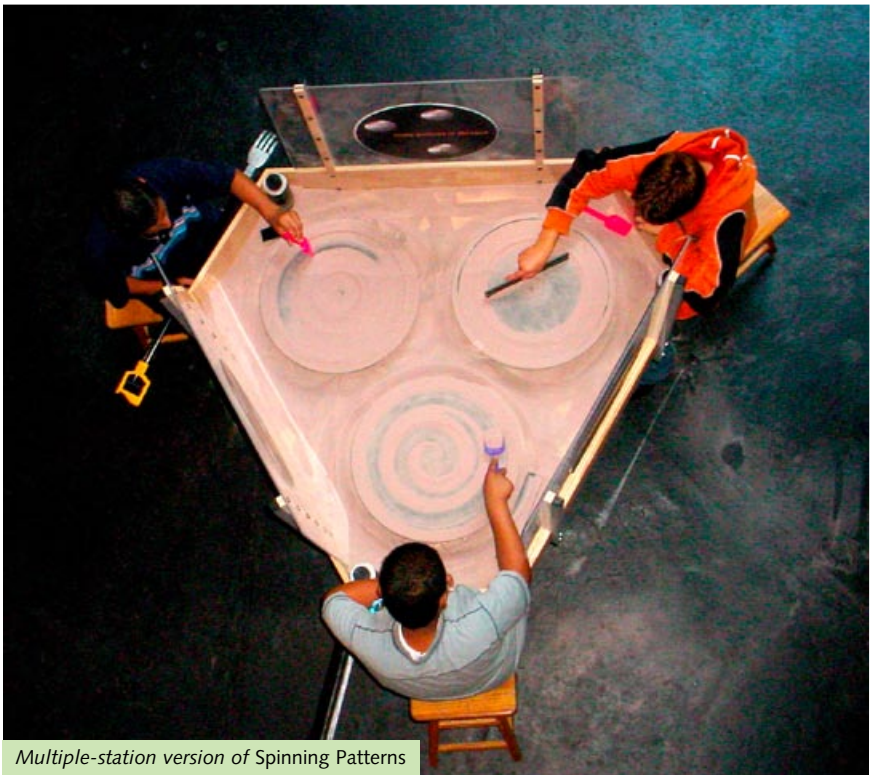
The holding time was longer: they spent twice as long at the multiple-station (7.7 minutes on average) as at the single-station version (3.4 minutes)[13].

The Physical Engagement score, judged by the complexity of the patterns they made, was higher at the multiple-station version[14].

The Intellectual Engagement score was marginally higher at the multiple-station version than at the single-station version[15], as measured by how intentionally visitors seemed to be drawing patterns and by whether they discussed techniques for drawing interesting shapes or letters. The main difference was that significantly more visitors to the multiple-station version talked about techniques for drawing patterns than did visitors to the single-station version (56% vs. 37%)[16].

The Social Engagement score indicated that 65% of the visitor groups at the single-station version interfered with one another while making patterns, while only 26% did so at the multiple-station version[17]. This underscored the need for individuals to have control over their workspace at the exhibit.

Single-station version of Spinning Patterns



Multiple-station version of Spinning Patterns

After watching more video footage of behavior at the three exhibits, we realized that interference between visitors at the single-station *Spinning Patterns* actually inhibited the creation and discussion of complex sand patterns. Essentially, visitor interference shut down engagement. But at the other two exhibits, interference at the single-station version was minor and did not cause serious problems for visitors, suggesting that the need for individual control at those exhibits was not as great. This suggests that *individual control is important when interference can disrupt the primary exhibit phenomenon.* The multiple-station approach seems to be helpful in ameliorating visitor interference while still providing opportunities for visitors to use the exhibit as a group. So the expectation that motivated the study—that multiple stations can help promote *APE*—was supported, but not in every case.

Summary

Taking all our results into account, it seems that the project was successful at building exhibits with the goal of promoting active prolonged engagement by visitors. Most compellingly, visitors at *APE* exhibits were more *actively* asking questions about using and understanding the exhibits, and they were answering their own questions rather than immediately turning to an exhibit graphic for explanations. Visitors were engaged for more *prolonged* periods of time at *APE* exhibits—often for more than four minutes. And they were engaged *differently* at *APE* exhibits, taking their own paths through the experience rather than following a fixed set of steps.

When we looked at the designs of *APE* exhibits, we found that visitors sensed they were both more open-ended than were *PD* exhibits and less likely to suggest an endpoint or stopping point to the exhibit activity. And our study of the multiple-station exhibit design showed that, when visitors interfered with one another at a single-station version, the multiple-station version was better at promoting *APE* behavior. This underscores the importance of designing exhibits so that visitors are able to use them in groups; multiple-station exhibits may indeed be a key solution to that design problem.

These findings, together with additional indicators of the success of *APE* exhibits at promoting active prolonged engagement, are summarized in Table 5.

Table 5. Evidence for *APE* Behavior

Attribute of *APE*	**Evidence for That Attribute in Visitor Behavior**
Visitors of different backgrounds are engaged together	We regularly saw intergenerational groups using *APE* exhibits.*
Visitors spend more time	Visitors spent 3 times longer at *APE* than at *PD* exhibits.
Visitors leave for extrinsic reasons	60% of visitors at *APE* exhibits left for extrinsic reasons, compared to 40% at *PD exhibits.*
Visitors ask and pursue their own questions	Visitors asked more questions that focused on using or understanding the exhibit at *APE* exhibits.
Visitors answer their own questions	Visitors answered questions by using or discussing the exhibit 77% of the time at *APE* exhibits but only 56% of the time at *PD* exhibits.
Visitors indicate they are practicing scientific inquiry skills	We regularly saw visitors engaged in the processes of observing, experimenting, explaining, and applying.*
Visitors continue using exhibits even after they read the exhibit graphic's explanation	We often saw visitors do this at *Investigation APE* exhibits, such as *Downhill Race* and *Gravity-Powered Calculator,* where the museum's explanation could most inhibit visitor-driven inquiry.
Visitors engage in activities that are not fully prescribed by the exhibit graphic	We often saw visitors engage in activities that were not fully dictated by the graphic.*
Visitors seem to be constructing a conceptual understanding	Visitors' conversations often indicated that they were making sense of the concepts underlying the *APE* exhibits.*

** Documented in formative evaluation reports and in Tisdal and Perry's summative evaluation reports.*

Notes

1. Several researchers have devised other ways either to gain visitors' consent for videotaping (e.g., Crowley and Callanan, 1998) or to avoid the issue entirely by restating visitors' utterances into a recorder (e.g., Borun et al., 1998). Unfortunately, none of these methods appeared practical in the Exploratorium's setting.
2. Our final method was approved by an institutional review board called Independent Review Consulting. It can be found on the World Wide Web at www.irb-irc.net.
3. Formative evaluation is conducted as exhibits are being developed in order to improve them.
4. Two researchers independently coded a random sample of 27 visitor groups. Interrater agreement on the code for each question was 81%; interrater agreement on the code for each response was 80%.
5. Mann-Whitney Test, $p = .70$.
6. This difference is statistically significant ($\chi^2 = 50.8$, $p < .001$).
7. This difference is statistically significant ($\chi^2 = 56.7$, $p < .001$).
8. This difference is statistically significant (Mann-Whitney Test, $p < .0001$).
9. This difference is statistically significant (Wilcoxon Signed Rank Test, $p < .0001$).
10. This study is described in detail in the first of two summative evaluation reports by Selinda Research Associates, entitled *Going APE! at the Exploratorium: An Interim Summative Evaluation Study*. The report is available on the Exploratorium's Visitor Research and Evaluation Web site: www.exploratorium.edu/partner/evaluation.
11. The six case studies from the follow-up report produced by Selinda Research Associates, *Phase 2 Summative Evaluation of Active Prolonged Engagement at the Exploratorium*, can also be found at the Exploratorium's Visitor Research and Evaluation Web site (see note 10, above).
12. This difference is statistically significant ($\chi^2 = 16.0$, $p < .001$).
13. This difference is statistically significant (Mann-Whitney Test, $p < .0001$).
14. This difference is statistically significant ($F_{1,97} = 5.2$, $p < .05$).
15. The difference is statistically marginal ($F_{1,97} = 7.6$, $p < .09$).
16. The difference is statistically marginal ($\chi^2 = 3.7$, $p < .06$).
17. This difference is statistically significant ($\chi^2 = 15.4$, $p < .0001$).