Learning Conversations in Museums

Edited by

Gaea Leinhardt
Kevin Crowley
Karen Knutson
University of Pittsburgh

LEA
2002
LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS
Mahwah, New Jersey London

Chapter 8

Looking for Learning in Visitor Talk: A Methodological Exploration

Sue Allen
The Exploratorium
San Francisco, CA

When the Exploratorium was asked to conduct one of the early studies to contribute to the Museum Learning Collaborative (Leinhardt & Crowley, 1998), I considered it a rare opportunity to contribute to both of the professional communities that I currently straddle.

As an educational researcher, I embraced the chance to study in-depth learning in the public space of the Exploratorium. One of my continuing goals at the museum is to explore and refine fruitful methods of collecting and analyzing evidence of learning in an environment that is highly challenging from a research perspective. In addition, I wanted to use the lens of visitor conversations to gather baseline research data about what visitors learn while visiting an Exploratorium exhibition. Few exhibitions are ever studied in this way, and the Exploratorium's recent temporary exhibitions are of particular theoretical interest because they combine elements from different museum traditions. For example, the *Frogs* exhibition that was the focus of this study contained hands-on interactive elements typical of a science museum, terrariums of live animals typical of a zoo or natural history museum, cases of cultural artifacts such as fetishes and musical instruments typical of a cultural history museum, and two-dimensional elements that could be read or looked at, such as maps and examples of froggy folklore. Such diversity is unusual in the museum field, and gave me an opportunity to make comparisons among the kinds of learning experiences visitors have with different types of exhibit elements.

As an exhibit evaluator active in the field of Visitor Studies, I also had an additional goal for the study. I wanted to explore possible methods for data gathering and analysis that might be of practical value to exhibit evaluators. There are both political and practical reasons for assessing the kinds of learning that happen in informal environments. In the face of increasing pressure to prove their effectiveness as educational institutions, more and more museums are conducting summative evaluations of their exhibitions. Such evaluations usually include tracking and timing studies of visitors' behavior (summarized in Serrell, 1998), often combined with a fairly tightly structured exit interview or questionnaire. It seems to me that these evaluations would benefit from including a component that assessed visitor learning through more naturalistic or open-ended study of visitor conversations, particularly those that happen "in real time" during the visit. In order to make this kind of study feasible, we need to develop methods that do not demand the resources of time, money, and research expertise that characterize large research efforts. To that end, this study had a pure research focus, but also explored some possible avenues for bringing a more sociocultural definition of learning to standard evaluation practices.

BACKGROUND

It is commonly accepted by museum evaluators and visitor studies professionals that school-based methods of assessing learning, such as conceptual pre- and posttests, do not transfer well to the study of learning in informal environments (e.g., Bitgood, Serrell, & Thompson, 1994; Crane, 1994; Falk & Dierking, 1992; Hein, 1998; Jeffery-Clay, 1998; Munley, 1992; Serrell, 1990). It is also well known that learning in museums is highly social in nature, and museum researchers are beginning to embrace sociocultural perspectives to describe learning (e.g., Crowley & Callanan, 1998; Falk & Dierking, 2000; Guberman & Van Dusen, 2001; Martin, 1996; Matusov & Rogoff, 1995; Silverman, 1990; Uzzell, 1993). However, most of the methods used to study visitors' experiences still rely on the responses of individuals rather than groups, and visitors' feedback is most often gleaned after they have left the exhibition.

One promising way of looking for the subtle, moment-by-moment learning that characterizes learning in museums is by analyzing visitors' conversations as they move through the public exhibit space. Although conversational analysis is not a routine part of most evaluations, a number of researchers have conducted such studies (e.g., Borun et al., 1998; Crowley & Callanan, 1998; Diamond, 1980; Guberman, Emo, Simmons, Taylor, & Sullivan, 1999; Hensel, 1987; Hilke, 1988; Lucas, McManus, & Thomas, 1986; McManus, 1988, 1989b; Silverman, 1990; Taylor, 1986; Tunnicliffe, 1998).

Of these, Silverman was unusual in focusing on exhibitions; other researchers have studied exhibit elements primarily, or have focused on an entire visit. I believe that the exhibition is a particularly interesting unit of study for several reasons. First, it is coherent conceptually, given that most exhibitions have a fairly well-defined and articulable scope and set of objectives; this gives a relatively clear curatorial framework with which to compare visitor experiences. Second, exhibitions are logistically suitable for conversational analysis, because they are large enough to show repeating patterns of visitor behaviors, but not so extensive as to require an entire day just to gather the data on a single family, as is often the case with whole-visit studies (e.g., Diamond, 1980). Lastly, exhibit development projects (such as those funded by federal agencies) usually generate exhibitions that require summative evaluations, so there is an opportunity to gather data that may serve both research and evaluation purposes.

Another key feature of the foregoing studies is their analysis schemes, which vary considerably. Stubbs (1983) pointed out that "discourse analysis is very difficult. We seem to be dealing with some kind of theory of social action" (p. 3). Indeed, I would argue that visitors' conversations as they move through an exhibition are particularly complex, because they involve such close ties between situation, knowledge, action, and language. Stubbs also notes that discourse analysis encompasses a wide variety of techniques and stances toward natural language, as researchers grapple with problems such as the multiple meanings that utterances can have. For example, he distinguishes between discourse acts, which are defined according to their function within the discourse itself (such as initiating, continuing, and terminating exchanges), and speech acts, which are defined according to psychological and social functions outside the ongoing discourse (such as naming, thanking, promising). From this perspective, the studies by McManus and Diamond focused more on discourse acts, such as "elicitation," "summons," "reply," "tell to read," and so forth. For the purposes of the *Frogs* study, I wanted to focus exclusively on speech acts, and in particular, those acts that functioned to advance the visitors' learning as defined next.

Whereas Hilke and Diamond devoted considerable attention to the differences in roles and relationships among members of a group, I preferred to make the simplifying assumption of treating all group members as collaborative learners, with a focus on the joint advancement of understanding. Also, based on my secondary goal of exploring practical discourse analysis methods for evaluators, I wanted to keep my utterance categories few in number and highly generalizable across different exhibitions. This was in contrast to Hilke and Tunnicliffe, each of whom devised a coding scheme with more than 70 categories of utterance, and in contrast to Tunnicliffe, whose analysis scheme involved specific content-specific categories at the more detailed levels (e.g., "mentions reproductive organs").

THEORETICAL FRAMEWORK AND RESEARCH QUESTIONS

My fundamental research goal was to characterize and quantify evidence of learning in the conversations of people visiting this exhibition. Further, I wanted to know whether "learning-talk" could be reliably identified as falling into different categories, and how common the various categories would be. Lastly, I wanted to characterize any differences between the patterns of learning-talk elicited by different types of exhibit elements, or between the conversations of groups that include and exclude children. Answers to all these questions would contribute to basic research on learning in informal environments, and would also inform museum practitioners who are in the business of designing learning experiences for diverse audiences, using a variety of exhibit-design strategies.

In defining "learning" for the purposes of this study, I used the following general framework:

1. I embraced the set of categories, generally attributed to Bloom (1956) which includes affective, cognitive, and pyschomotor learning. These categories are frequently used by museum studies professionals because they map well onto the kinds of experiences visitors usually have in museums: thinking, feeling, and interacting with objects.

2. I took from the sociocultural perspective a focus on learning as an interpretive act of meaning-making, a process rather than an outcome, and a joint activity of a group, rather than being attributable to one of the people only. For this reason, I did not attempt to analyze the learning of individual visitors, but rather to characterize learning by the group.

3. I wanted to be sure that my definition was not so tied to formal learning assessments that it would exclude the kinds of learning that are most prevalent in informal settings. In particular, research has shown that museum learning tends to be affective, personal, sporadic, concrete rather than abstract, and associational rather than deductive. Even explanations tend to be brief, partial, and nonhierarchical (e.g., Crowley, Galco, Jacobs, & Russo, 2000).

4. In terms of the scope of the content, I wanted to define *learning* quite narrowly to refer to discussion of the exhibits and the exhibition, and its topic area. In other words, I excluded learning-talk that related to other parts of the museum visit. I also excluded navigational talk such as beckoning over, or noting that an exhibit has been missed. This choice was made partly for political reasons, to respond to the recurring comment, "Yes, visitors have fun in museums, but what do they really learn?" By starting with a definition of learning that constrains content, I hoped to put the museum in a stronger position to make claims about its efficacy as a learning institution.

My definition of learning did not require intentionality on the part of the speaker or listener (which would have been very difficult to infer, and is also not a necessary condition for advancing meaning-making). Also, learning did not require explicit expression of awareness or acknowledgment, although such statements did end up being coded as a subcategory (metacognition) of the learning-talk.

Throughout the study, my bottom-line question was always: Is this evidence of learning? From a cognitive science perspective: Is it likely that one or both of these people have just acquired new knowledge or ability from what was said? Or, from a more sociocultural perspective: Has this utterance advanced the dyad's collaborative process of making meaning from the exhibition? If so, I tried to include it somewhere in the coding framework.

OVERVIEW OF THE *FROGS* EXHIBITION

The context for the learning study was *Frogs*, a temporary exhibition built by Exploratorium staff in collaboration with several consultants who had expertise in amphibians, exhibition design, and exhibitions involving live animals. It was open from February 1999 to February 2000, and occupied approximately 4,000 square feet of museum floor space.

Front-end evaluation studies conducted early in the development process revealed that Exploratorium visitors thought frogs to be an interesting topic, but most knew little about them. Many visitors reported having had some contact with amphibians from high school biology classes, and had listened to frogs calling on spring or summer evenings. Some visitors knew about malformed frogs.

Informed by these data, the development team created the following goals and conceptual outline: (a) to present scientific, social and cultural aspects of people's relationship to frogs, (b) to engender respect and appreciation for the animals in the museum's visitors, (c) to create something that would be beautiful, intriguing, and informative for the museum's age-diverse audience.

In terms of content subdivisions, the exhibition included: an introductory area that defined frogs and toads and amphibian development, sections on eating and being eaten, frog and toad calls, a showcasing of amphibian anatomy, a close-up observation area, a cluster of exhibits showing "amazing adaptations," an area that discussed the declining status of frogs worldwide, and a section on frog locomotion.

As already mentioned, the final exhibition contained an unusual diversity of different *types* of exhibit elements, in an attempt to illustrate the broad variety of ways humans have understood and connected with frogs, as well as to engage visitors with different learning styles. In all, there were:

- 10 hands-on interactive elements typical of a science museum,
- 23 terrariums of live frogs and toads, typical of a zoo or natural history museum,
- 5 cases of cultural artifacts such as fetishes, baskets and musical instruments typical of a cultural history museum,
- 18 two-dimensional elements that could be read or looked at, such as maps, excerpts from children's books, and renderings of froggy folklore from many cultures,
- 3 cases of organic material, such as preserved frogs or frog food,
- 3 videos of frog activities; 2 windows to the "Frog Lab" where frogs were allowed to rest and breed,
- an extended entry bridge, and
- an immersion experience of sitting on a back porch at night and listening to a chorus of frogs.

METHODOLOGICAL CHALLENGES

Studying conversations, especially at a highly interactive museum such as the Exploratorium, is a logistically difficult and expensive undertaking. The environment of the museum floor presents a set of challenges beyond those inherent in most laboratory or school settings.

First, the acoustics of the public spaces are very poor, and there is a huge amount of ambient noise. Major sources of sound include background hubbub from distant visitors, screams from excited children, conversations by nearby visitors (easy to mistake for members of the study group), and the myriad of sounds that interactive exhibits make when in use, such as bangs, dings, music, ratcheting, splashes, music, and prerecorded speech.

Second, visitors move around a great deal, and the groups they arrive in keep changing and reforming on short time-scales. This means that microphones in static locations can only catch conversations over small intervals of time (typically a minute), and researchers wanting to record extended conversations have to use cordless microphones, with their potential for interference and other audio problems. The fluid movements of visiting groups also complicate the decision of which visitors should be asked to wear a microphone, because the parties to a conversation keep changing, so that different conversations might take place simultaneously among different members of a group.

Another logistical difficulty is that much of the visitor activity involves interacting with the specific objects in the environment, so it is almost impossible to make sense of an extended audio stream without corresponding video or observation data.

Once the conversations have been recorded, transcription is expensive and time-consuming, and the recordings are particularly challenging to decipher because visitors speak with a great variety of accents and styles, and bring a huge diversity of previous experiences that might be referred to at any time.

Even if visitors' words are audible, coding of the conversations is difficult because much of the discourse is fragmented, ambiguous, or lacking clear referents. One reason for this is that in conversation, speakers seldom articulate what is obvious to listeners: the details of the physical situation they are in, what they are attending to, the set of common experiences they have shared over the years, the "short-hand" words and phrases they have come to use together. As Garvey (1984) put it, "in conversation . . . if something can be taken for granted, it usually is" (p. 12).

Finally, interpretation of the results is especially challenging because the museum environment is dense and complex, with many variables that can influence what visitors say and do not say. For example, conversations are likely to depend on visitor variables (such as demographics, psychographics, previous experiences, interests, attitudes, expectations, group dynamics, and current state of comfort and energy), exhibition variables (such as location within the museum, degree of orientation provided, lighting, seating, ambient noise, and current crowdedness), and variables in the detailed design of individual exhibit elements (such as height, coloration, physical accessibility, interface, display style, label content and tone). It is seldom easy to draw conclusions about what kinds of exhibit environments lead to what kinds of conversations.

Addressing the Audio Challenge

Recording extended visitor conversations has been historically difficult to achieve in the noisy, open, fluid environment of a hands-on science museum (e.g., Borun et al., 1998). However, recent advances in audio technology have made it possible to buy high-quality cordless radio microphones at a reasonable cost. We used two such microphones, putting them in "fanny packs" which the visitors could wear around their waists. At a convenient remote location (just behind a high wall bordering the public space), we mixed the signals together and recorded the conversations onto regular audio cassettes. We also made backup cassette recordings of the components of the conversation, directly from the receiver and before mixing.

The quality of the sound was generally high, and limited mainly by the degree of clarity of the speaker. Only one tape was rejected from the study because of inadequate audio quality, and this was a tape of visitors who were speaking extremely quietly to each other, almost as if they were whispering. Given the extremely high level of ambient noise in the Exploratorium, and the high degree of crowdedness of the exhibition, we were very satisfied with the audio quality achieved.

Gathering Synchronous Movement Data

We wanted to know where visitors were when they were speaking, so that we could relate their conversations to the 66 specific exhibit elements comprising the *Frogs* exhibition. We considered videotaping to get the most precise information about visitors' movements, but eventually dismissed this as unfeasible, mostly because it would have taken a large number of fixed cameras to cover the 4,000 square-foot exhibition, with its many twists and turns. A tracker carrying a camera would have had to keep very close to the visitors in order to get good enough video to code, which we thought would be highly intrusive. Furthermore, it would have been very difficult to maintain a video stream that is at once steady enough to be useful, and fast-moving enough to record the positions and behaviors of two people who may be moving somewhat independently of each other as they explore the exhibition.

Instead, we obtained information on visitors' locations by having a "tracker" follow them discreetly, noting their movements and behaviors. In order to get synchronicity of this information with the conversations, we used a system of three microphones: two worn by the visitors, and one by the tracker. The tracker's audio stream was recorded onto the Right channel of a tape, while the mixed signal from the two visitors was simultaneously recorded on the Left channel. The result is a set of tapes that contain accurate information about where visitors were when they spoke, to an accuracy of a few seconds, assuming the tracker was alert to all movements. This technique proved to be a successful way of keeping the two audio streams synchronous, although (as described later) we ended up tolerating some looseness in the timing of the tracking call-outs.

Visitor Selection

Based on our pilot observations of visitor groups moving through the *Frogs* exhibition, we decided to conduct the study using only groups of two visitors (dyads). Part of the reason for this was our concern that the tracker would not be able to monitor more than two people simultaneously. In addition, we noted that larger groups tended to split up and reform in different combinations, yielding the possibility of different simultaneous conversations, which would result in a garbled sound on the recording. During the analysis phase, we discovered yet another reason for excluding larger groups. Because they frequently split up and reform, the visitors in these groups are much more likely to talk or ask each other about what they have

previously seen or done at other exhibit elements. Such talk adds to the complexity of the coding scheme because it introduces new categories of talk that are about the exhibits but have a reflective aspect (e.g., descriptions of exhibit features remembered but not currently visible).

We thus simplified the study by inviting only dyads to participate. Our complete list of criteria for participants was as follows:

1. The visitors were in a party of two.
2. They spoke English as their first language, ensuring that their conversations would be understandable to the coders, but also normal and comfortable for them.
3. It was their first visit to the *Frogs* exhibition, so that we would be studying conversations about real-time experiences rather than the more complex case of real-time experiences combined with reflections.
4. Both visitors were 18 or older, or a child[1] was accompanied by a parent or guardian, so that they could give informed consent to be recorded.

Of the 118 dyads we approached, 45 (38%) declined our invitation to participate. Most of these gave no specific reason for their decision, and we did not ask them to justify it. The most common reasons that were spontaneously given were lack of time (6), other arranged meetings within the museum (6), or the difficulty of making it work with a fast-moving, young, or reluctant child (12). We ourselves refused a further 24 dyads (20%) who expressed initial interest, but proved to be ineligible to participate. Of these, the largest group (14) were visitors who had already seen the exhibition. We decided to exclude this group for simplicity, to avoid the complications of having a subgroup of visitors who might only re-visit parts of the exhibition, and whose talk would be an ambiguous combination of real-time experience and reflection. The remaining 49 dyads (42%) were both eligible and willing to participate in the study.

We felt that the use of dyads was an excellent way to study conversations and keep the complexity of the situation at manageable levels. The only major drawback to this method was the amount of time it took to recruit participants. Our constraints on group size and age meant that we spent relatively long periods simply waiting for an appropriate group to approach, and this was sometimes disheartening for a recruiter. Even with an acceptance rate that was much higher than expected, it took us about 20 days of data collection to record and interview 49 visitor dyads. We found that we could collect data from an average of 3 to 4 dyads on a given weekend day, and 0 to 2

[1]We experimented with the minimum age of child participants. We succeeded in recording (and later interviewing) children as young as four.

on a weekday afternoon, once the school fieldtrip groups had left the museum. It was unfortunate that the exhibition was near the rear of the Exploratorium, so that it took about an hour for visitors to begin arriving there. We also found that we could not recruit any participants in the last 90 minutes of the museum's open hours, because visitors expressed anxiety about getting through the exhibition in time to see other things before closing time. These constraints left a productive period of perhaps 4½ hours each weekend day, and 2 hours each weekday.

To recruit a higher number of participants, we tried letting larger family groups participate by designating two people to carry the microphones. This met with limited success. Some families did act on our requests to go through the exhibition in two groups (the dyad separate from the rest) without all talking together, but in most cases the excitement of the exhibition was too much for them not to share what they were seeing with family members on the other side of the gallery. This created serious problems in the transcription, because family members who mingled in this way seemed even more likely to compare reflections on what they had already seen and done, which complicated later coding considerably. We recommend using this technique only if data-collection time is severely restricted, and even then, suggesting strongly to the family that the two groups look at different parts of the museum while the study is underway.

Obtaining Informed Consent

When recruiting visitors at the entry to the museum, we told them that we would be recording their conversations so that the museum staff could learn more about visitors' experiences in the exhibition. We also told them that we would be watching them to see where they went within the exhibition. We did not ask them for written consent, however, until they had finished seeing the exhibition, and after the interview had taken place. There were two reasons for delaying the signing of consent forms. First, we found that visitors were much more relaxed at the end of seeing the exhibition than at the beginning, and took time to consider what we were saying or to ask questions. The recruitment point, being at the most inviting entry point to the exhibition, was a place where visitors tended to be focused on getting into the exhibition, especially when the dyads included children.

Second, at the end of their visit visitors were in a better position to assess how comfortable they were with what they had actually revealed during their conversations. Had anyone refused to give their consent at that point (and nobody did), we would have given them their audiotape to take home with them. We also offered all dyads a choice in whether to let their tape be available for use at conferences and larger professional settings beyond the museum, and one dyad decided not to allow this.

Tracking Visitors

We had already told visitors that we would be following them to see where they stopped. Nevertheless, we felt it was important to track them unobtrusively, so as to influence their behavior as little as possible. The tracker dressed in dark colors so that her own cordless microphone (pinned so as to be near her mouth) would be inconspicuous, and she spoke in low tones as she followed behind the dyad, concealing herself behind exhibit elements where possible. The only times this technique proved ineffective were on the rare occasions when the exhibition was empty; in one such case, a visitor being tracked actually approached the tracker and attempted to start a conversation. The best solution to this problem, if staffing allows, is for the person who recruits to be different from the tracker; this minimizes the chance of the visitors becoming conscious of the tracker while moving through the exhibition.

The most important thing the tracker did was to note where the two visitors were at all times, irrespective of whether they were talking or not. The tracker's audio stream went something like this: "She stops at Jungle . . . he joins, he moves on . . . to . . . Spadefoot . . . she joins . . ." We defined a "stop" as happening whenever someone's two feet come to a halt in front of an element, and their eyes are on the element. Unlike the case in many tracking studies (e.g., Serrell, 1998), we did not wait for a minimum of several seconds before counting something as a stop; reporting two simultaneous movements was too taxing for any further judgments about minimum time, and in any case, we wanted to include everything they saw (however briefly) in case they talked about it later.

One key behavior we overlooked was interactions with other visitors. When the museum is busy, many people may be crowded around an exhibit element, and some of the audiotapes include voices that we can't unambiguously identify as belonging to our tracked visitors. For this reason, we would recommend for future studies that the tracker attempt to record the presence of nearby visitors, and, in particular, the moments when they seem to be talking.

How Authentic Are the Visitors' Conversations?

One of the main concerns that museum professionals have about microphoning visitors is that this may affect the nature of the talk and behavior in unpredictable ways. Although a full assessment of such impacts was beyond the scope of this study, we did make some attempts to assess the most obvious kinds of impact.

When coding the visitors' transcripts, we created a code for verbal evidence that visitors were aware of the microphones or the fact that they were acting as research subjects. Of the 30 dyads whose conversations were transcribed in detail, 9 (30%) mentioned something about the research as they first entered the exhibition. These initial comments included expressions of pleasure and support for the research, suggestions about dealing with the microphones, and jokes about talking in code or not swearing.

Once they had entered the exhibition, 14 of the 30 dyads (47%) made some reference to the microphones or audiotaping. Most commonly, parents encouraged children to keep their microphones on but not to play with them (6 dyads), individuals talked about removing their microphones to participate in a jumping activity (6 dyads), and dyads who were visiting with a larger group explained what they were doing (6 dyads). Although such comments were more frequent than we had hoped, in most cases it was a single incident, took no more than a few seconds, and did not result in any obvious derailment of activity (such as an explicit decision to move somewhere or do something). The majority of dyads never mentioned the research or the microphones at all.

We attempted to minimize visitors' awareness of the recordings by keeping out of their view as much as possible and by using fanny packs that they could "strap on and forget." However, there is a limit to how unobtrusive we can be, because our ethical policies require that we tell visitors what we are about to do, and ask for their consent to be recorded. During the pilot phase of the study, we also experimented with different recruiting locations in an attempt to lessen the impact of visitor awareness, but this strategy proved to be laughably unsuccessful.[2]

There was at least one dimension of visitor behavior that was significantly affected by the methods used in our study. The dyads who participated stayed much longer in the exhibition than visitors who were unobtrusively tracked. A study of the *Frogs* exhibition by Contini (1999) showed that the average time spent in the exhibition was 10.9 minutes, and the median time 9.0 minutes. By contrast, the average time spent by visitors in the conversation study was 25.4 minutes, and the median time was 24.1 minutes (more than twice as long). Even so, the longer times do not necessarily imply that the visitors in the study were behaving unusually. At least part of the time difference may be due to a selection effect. Visitors who chose (for a variety of possible reasons) to take a "quick look" at the exhibition would have been included in the tracking study, but would probably have refused to participate in an extended study. Moreover, McManus (1987) found that the constituency characterized by the briefest stay at exhibits at the British Museum (Natural History) was the "singletons" (groups of one), and this constituency was entirely excluded from our conversation study.

We take the view that every powerful research method yields insights and creates distortions, so that we would use this kind of method in combination with others that complement it. For example, the standard tracking methods that are common to museum evaluations give a much better sense of how long visitors stay in exhibitions, where they stop, and for how long; we would not propose to compare our timing data directly with that collected unobtrusively, nor would we calculate quantities such as "sweep rates" or "percentage diligent visitors" (Serrell, 1998). However, standard tracking methods leave us wondering about visitors' interpretations of what they are seeing, and how they think about it, moment by moment. In understanding these kinds of learning processes, analysis of conversations is an appropriate and powerful tool.

Perhaps it is most appropriate to compare recording visitors to conducting a "cued interview," one of the more common techniques in visitor studies. It has the same qualities of alerting visitors to the fact that they are under scrutiny, so one might expect them to pay closer attention to the exhibits, to try harder, talk more, or generally be "on their best behavior." For this reason, we cautiously take the position of regarding the conversations as a probable "best case" display of visitors' authentic behavior.

[2]We hoped that we could find a way to make visitors less self-conscious by giving them the microphones without them realizing that the *Frogs* exhibition was the focus of our interest. We anticipated some kind of exponential drop-off in visitors' self-consciousness, so we reasoned that the ideal situation would be for the visitors to forget about the microphones just around the time they happened to enter the *Frogs* exhibition. We explored this option by asking 20 groups of visitors who were entering *Frogs*, where they had been about 20 minutes previously. We then recruited visitors at the point that seemed the most popular "upstream" location, and gave them microphones, telling them that we would catch up with them at a later time, probably about 45 minutes hence. This technique proved disastrous, when visitors happily moved in tantalizingly different directions from the *Frogs* entrance, and we were forced to sit in the audio room, waiting impatiently and hoping for our true study to begin. The experiment ended abruptly, and we resigned ourselves to recruiting at the entrance to the exhibition for the sake of efficiency.

Transcription

One of the most time-consuming aspects of analyzing the conversations was getting them transcribed. With a modest transcription budget, we were only able to get 15 tapes transcribed professionally. Thereafter, we put a call out to museum volunteers, who were enthusiastic but limited in the amount of time they could commit to the project. We ended the study with a total of 30 transcripts.

On the whole we were disappointed with the quality of the transcriptions, both professional and done by volunteers. Even after sampling a range of different companies, we found that we always had to review the entire tape and make many edits. We found that we could hear almost everything the visitors were saying, but the transcribers obviously struggled.[3] We suspect that this difference was due to the transcribers' complete lack of familiarity with the exhibition. We did notice an increase in our own ability to hear visitors after we printed out the entire label text of the exhibition and became very familiar with its contents as part of the coding process, so we conclude that familiarity with the exhibition is a key requirement for good transcription.

Suggestions for Obtaining Intelligible Audio Recordings Without Video

At this point I wish to respond to the challenges raised at the beginning of this chapter, by summarizing our solution for obtaining intelligible audio recordings without video, even in a very active and noisy public space:

- Get a set of at least three high-quality cordless microphones, and test that they give a strong signal over all parts of the museum.
- Recruit visitors who arrive in groups of two, speak English as their first language, and are at least 4 years old. The reduction in complexity will be worth the extra wait.
- Document the exhibition in fine detail so that you can refer to labels and exhibits later, as questions arise in the analysis.
- Employ a transcriber who is intimately familiar with the exhibition.

With all of these in place, we felt able to make sense of almost everything visitors said in the tapes, without the need for video. We should say, however, that one of the reasons for our success may be that the commonest type of exhibit in the *Frogs* exhibition was the terrarium containing live animals, rather than the hands-on interactive exhibits for which the Exploratorium is well known. It may be that visitors were forced to communicate with each other using language, in part because there were reduced opportunities for them to communicate via more physical means.

Coding Visitors' Locations

We were pleased to discover, on listening to the Right and Left channels of the tapes, that most of the time visitors' talk was about the exhibit that was currently in front of them. Most of the exceptions were due to a "movement

[3]For example, "What frogs and toads eat, insects" was transcribed as "Can frogs and toads mate? It's sex."

lag," during which visitors would notice another exhibit, walk toward it, and start to comment on it before they came to a stop in front of it (e.g., a visitor might say, "Wow, look, orange frogs! Aren't they gorgeous" before coming to a stop). Because we wanted to know what kinds of learning were *evoked* by what types of exhibit elements, we chose to assign such statements to the exhibit that visitors were focusing on, rather than the one they were about to leave. Thus, our coding for location was based primarily on the exhibit element that visitors were talking *about*, but informed by the tracking information, which told us where they were when they were static, and which exhibits they were moving between.

For this step in the coding (parsing the transcripts based on exhibit element discussed), the straightforwardness of the task was reflected in our high intercoder reliability. We measured a 100% level of intercoder agreement about which exhibit elements were discussed, and a 90% level of agreement about which exact word in the transcript signified the start of the next exhibit element. In a similar study, Silverman (1990) also reported ease with this part of the coding: "With the aid of observation notes indicating where in the exhibit the pair was, as well as the pair's own comments, conversations were fairly easily broken down into object-related interactions for coding."

We also used the tracking information to create a list of exhibits that the visitors stopped at, but did not speak about ("silent stops"). We did this because we wanted to know all the exhibits visitors had seen, partly to get a measure of how common talking was compared to stopping, and partly to inform us of their total experience when it was time to do the exit interviews.

To our surprise, we found that we did not need a code for visitors "between exhibits" because it was very rare for conversations to happen between exhibits that were neither about the exhibit just left nor the one being approached. Perhaps it was partly due to visitors' awareness of the microphones, or partly due to the strength of the exhibition, but we found that visitors' conversations were tightly coupled with what they were seeing; so much so, that we were often able to guess movement information for the bulk of a written transcript, prior to hearing the tracking audio channel. The only times we ran into the problem of visitors talking about exhibits they were not near was with dyads that were part of larger families. In two such cases, the larger group met up with our dyad in the course of visiting the exhibition, and began talking with each other about what they had seen and done at remote exhibits. This behavior confirmed our preference for recruiting only true dyads in future studies, so that such family retrospectives would not arise. We found that when true dyads split up and rejoined, they tended to beckon each other over to see what was of interest, rather than talking about it later in the exhibition.

Creating a Specific Coding Scheme

In creating a framework by which to code the conversations, I combined approaches from sociocultural and cognitive science perspectives. I took a sociocultural approach when we decided not to categorize the conversations by individual speaker, but instead took verbal expressions of noticing, thinking, feeling, and acting, all as evidence that learning was taking place within the pair of companions. On the other hand, the specific categories of talk I used were mostly based on cognitive concepts such as attention, memory, declarative knowledge, inference, planning, and metacognition. I also included the categories of "Affective" and "Strategic," both of which have generally been regarded as fundamental areas of strength in informal learning environments, and which extend the assessment of learning beyond the strictly cognitive realm.

Within the basic framework already outlined, I tried to let the details of the categories and subcategories be "emergent," that is, shaped by the nature of the conversations themselves. To do this, coder Marni Goldman and I had to iterate on the coding scheme many times until we reached a stable and reliable set of categories that seemed to span the space of all types of learning talk. Ultimately, our coding system consisted of 5 categories and 16 subcategories of "learning-talk," utterances that we took as evidence of learning.

The Coding Scheme in Detail

Figure 8.1 shows the hierarchical structure of the final coding scheme, with 5 main categories and 16 subcategories of learning-talk.

1. Perceptual Talk. This category included all kinds of talk that had to do with visitors drawing attention to something in the sea of stimulus surrounding them. We regard it as evidence of learning because it is an act of identifying and sharing what is significant in a complex environment. The category included four subdivisions:

(a) Identification = pointing out something to attend to, such as an object or interesting part of the exhibit (e.g., "Oh, look at this guy," or "There's a tube.")

(b) Naming = stating the name of an object in the exhibit (e.g., "Oh, it's a Golden Frog.")

(c) Feature = pointing out some concrete aspect or property of the exhibit (e.g., "Check out the bump on his head," or "That's loud, huh?")

(d) Quotation = drawing attention to exhibit text by reading aloud part of a label. Needs to be an exact quote or a very close paraphrase (also known as "text echo," McManus, 1989a), (e.g., "One of the most common frogs in North America, and a typical jumping frog.")

Perceptual

Identification
Naming
Feature
Quotation

Conceptual

Simple
Complex
Prediction
Metacognition

Connecting

Life-connection
Knowledge-connection
Inter-exhibit connection

Strategic

Use
Metaperformance

Affective

Pleasure
Displeasure
Intrigue / Surprise

FIG. 8.1. Coding scheme, with 5 categories and 16 subcategories of learning-talk.

2. Conceptual Talk. This category captured cognitive interpretations of whatever was being attended to in the exhibit. To classify as a "cognitive interpretation," an utterance did not necessarily have to be abstract, have multiple steps, or reach a profound conclusion. We wanted our coding scheme to capture the breadth of small, individual, literal inferences that seemed far more typical of the conversations elicited by exhibit elements. The subcategories of conceptual talk were:

(a) Simple inference = single interpretive statement or interpretation of part of an exhibit (e.g., "They eat mice" after seeing a jar containing a mouse in a display of frog food, or "See now it looks like it's swimming" after successfully using a zoetrope to simulate the motion of a swimming frog)

(b) Complex inference = any hypothesis, generalization of exhibit information, or statement that discusses relationship between objects or properties (e.g., "That's a lot of body for them skinny legs to carry around," or "That would be hard to carve, wouldn't it?") Although we used the term *complex* to distinguish this from the very simplest kind of inference, it should be borne in mind that these were often single words or phrases rather than a formal series of deductions. The key criterion was that visitors draw some kind of inference about the exhibit element, beyond correctly interpreting what has been explicitly displayed.

(c) Prediction = stated expectation of what will happen, including what visitors are about to see or do (e.g., "I think it's going to be 'kwaa kwaa

kwaa' " when anticipating the sound of a particular frog call, or "Yeah, you'll start to grow legs" when viewing a tadpole.)

(d) Metacognition = reflection on one's own state of current or previous knowledge (e.g., "I didn't realize they could get that big," or "I can't remember, but I recognize him.")

3. Connecting Talk. This included any kind of talk that made explicit connections between something in the exhibition and some other knowledge or experience beyond it. Although we assume that all learning-talk involves previous knowledge to some degree, we felt that some types of utterance were distinct in that visitors were using the exhibits as a stimulus to share a personal story or previously learned information that was not directly coupled to what they were looking at. Subcategories were:

(a) Life connection = story, personal association, or likening of exhibit element to something familiar (e.g., "Yeah, my grandmother loves to collect stuff with frogs all over it" or "It looks like a brick, a floating brick.")

(b) Knowledge connection = declaration of knowledge gained prior to visiting the exhibition (e.g., "In Florida the dogs eat poisonous toads and die" or "Frogs, when it lays their egg, their egg floats to the top.")

(c) Inter-exhibit connection = any kind of link between exhibit elements, including the bringing of information gleaned at a previously visited element to the discussion of the current element. (e.g., "That's what I said. It eats anything as long as it fits in its mouth" when referring to the label from a previous part of the exhibition.)

4. Strategic Talk. Strategic talk (talk about strategies) was explicit discussion of how to use exhibits. It was not limited to hands-on exhibit elements, but was defined so as to include descriptions of how to move, where to look, or how to listen to something. Subcategories were:

(a) Use = statements about how to use an exhibit (e.g., "You're supposed to play this like that" when playing a wooden instrument that sounds like a frog, or "Okay go down to the water . . . and then go towards the back. See that little leafy type thing?" when searching for the leaf frog.)

(b) Metaperformance = expressions of evaluation of one's own or partner's performance, actions, or abilities (e.g., "I don't think I did a very good job of it.")

5. Affective Talk. In this category, we tried to capture all expressions of feeling, including pleasure, displeasure, and surprise or intrigue. Subcategories were:

(a) Pleasure = expressions of positive feelings or appreciation of aspects of an exhibit (e.g., "beautiful," "wonderful," "cool," "I like that one,") This subcategory also included laughter.

(b) Displeasure = expressions of negative feelings or dislike towards aspects of an exhibit, including sadness or sympathy (e.g., "poor thing," "ugly," "eeew," "gross," "yuck.")

(c) Intrigue = expressions of fascination or surprise (e.g., "wow," "gosh," "woah," "ooooh.")

Two caveats are worth mentioning with respect to the *affective talk* category. First, these verbal measures are, at best, crude indicators of the affective impact of an exhibition. However, we think it is worth making an attempt to capture affective learning-talk as best we can, and the prevalence in the conversations of words like those listed previously does suggest that they carry some significant aspect of visitors' shared experience. (A helpful review of some other approaches to affect is given in Roberts, 1990). Second, expressions of displeasure by visitors are not necessarily criticisms of the exhibition. In the case of the *Frogs* exhibition, some of the displays of organic material were deliberately graphic, and many children responded to videos of frogs eating such things as maggots with cries of "Ewwww!" yet seemed highly engaged and almost delighted.

Last, we note that the coding scheme we developed does depend quite sensitively on the detailed content of the exhibition, and the labels in particular. A single statement, such as "These are declining in California" could be a quote, a simple inference, a complex inference, a knowledge connection, or even an inter-exhibit connection, depending on what is written in the label-text at that exhibit element.

Counting Instances of Learning-Talk

As described earlier, we broke each conversation into segments of speech based on which exhibit element was being discussed. For each block, we then coded for the presence or absence of each of the 16 subcategories of learning-talk. This resulted in a coding matrix of 1's and 0's for each dyad, covering all types of learning-talk at all exhibit elements stopped at.

We did explore the possibility of coding the *frequency* of each kind of learning-talk for a given exhibit element, but decided against it based on the difficulty of determining what constituted different cases of a certain type of talk. Visitors' conversations often included repetitions, with slight variations or elaborations of each other's comments; these would have been quite difficult to count. For example:

Adult: Oh, I like them, these ones I like. Look at them.

Child: Oh, there's a whole bunch!
Adult: These are from Texas.
Child: Mom there's a thousand of them in there!
Adult: Those are from Texas, they're Texas ones. They look dry.
Child: They're everywhere, look.
Adult: And if they're from dry land, would they be frogs or Toads, John?
Child: Toads.
Adult: If they're from Texas, would they be frogs or toads?
Child: Toads.
Adult: Toads, yeah.

This kind of echoing was especially prevalent in the conversations of child–adult dyads, but was also quite common among adult–adult dyads. It made counting frequencies extremely difficult, and led to our decision to code only for presence or absence of each type of talk for each exhibit element. In spite of its coarseness, our "digital" approach nevertheless yielded a detailed profile of each dyad's conversation, owing to the large size of the coding matrix, and the fact that visitors tended to talk relatively briefly at a relatively large number of exhibit elements. We note that Silverman (1990) coded her conversations the same way, although she did not elaborate on the reasons behind this choice.

In combining the different subcategories of talk into the five larger categories, we used the logical "OR" function for each individual dyad. In other words, we coded any given dyad as having expressed "strategic learning-talk" at a given exhibit element if their conversation included at least one example of any of the subdivisions of strategic talk: use-talk, metaperformance talk, or both.

Intercoder Reliability

The coding process consisted of two steps: (a) parsing each transcript into segments of talk, one for each of the exhibit elements talked about, and (b) for each exhibit element, deciding which of the 16 types of talk were present. Because the two steps are not independent, we assessed intercoder reliability for both steps together.[4] Overall, we found our coding to be 78% reliable. We noted that this figure rose to 84% if we let go of the distinction between the subtly different categories of *simple* and *complex* cognitive talk.

[4]In the matrix of exhibit elements by talk subcategories, we ignored all cells that both coders assigned a "0" to; in other words, the large number of "absent" types of talk were not counted as agreements. We then counted as agreements all cells which both coders assigned "1" (i.e., cases where both agreed that a certain subcategory of talk was present in the conversation of a particular dyad at a particular exhibit element.) To determine the inter-coder reliability, we divided this number of agreements by the total of agreements plus disagreements (the Jaccard Coefficient of similarity). In counting disagreements, we included every mismatch in every cell of the coding matrix, even though this meant that a single coding of a single utterance by two coders could result in as many as two disagreements in the count, but could lead to one agreement at most. For this reason, we consider our reliability measure to be somewhat conservative.

Specific Questions Driving Conversational Analysis

As outlined earlier, my principal goal for this study was to characterize the kinds and frequencies of learning, as revealed by visitors' conversations in the *Frogs* exhibition. In addition, the multifaceted nature of the exhibition gave me an opportunity to use these conversational assessments to compare the learning patterns of visitors at different types of exhibit element. More specifically, I had the following questions regarding the conversation data:

- What was the average frequency of learning-talk, albeit in a situation where visitors were cued? Was it something generally rare or common?
- What categories and subcategories of learning-talk emerged from the data, and what were their relative frequencies?
- Did different types of elements within the exhibition (e.g., hands-on, live, readable) tend to elicit different kinds of learning-talk?
- Were there any differences between the learning-talk of adult–adult (AA) and child–adult (CA) dyads, either in frequency or type?
- Given that one of the goals of the exhibition was to engender respect and appreciation for frogs, did visitors express any intention to take personal action or change their behaviors, based on the exhibition?

RESULTS

Here I present the key results of the conversational analysis, beginning with the number of stops made by dyads at the various elements in the exhibition. The data on stopping are important because they were used as a normalization factor in subsequent analyses. In other words, every frequency of learning-talk was calculated as a percentage of the elements where each dyad chose to stop, because I regarded these "stopped" elements as the only ones providing real opportunities for learning conversations to occur.

Number of Stops

On average, the visitor dyads stopped[5] at 34 exhibit elements (51% of the total on display), and spent an average of 25.4 minutes in the exhibition.

[5]An element was considered to be a stop if either one of the visitors in a dyad, or both, stopped there.

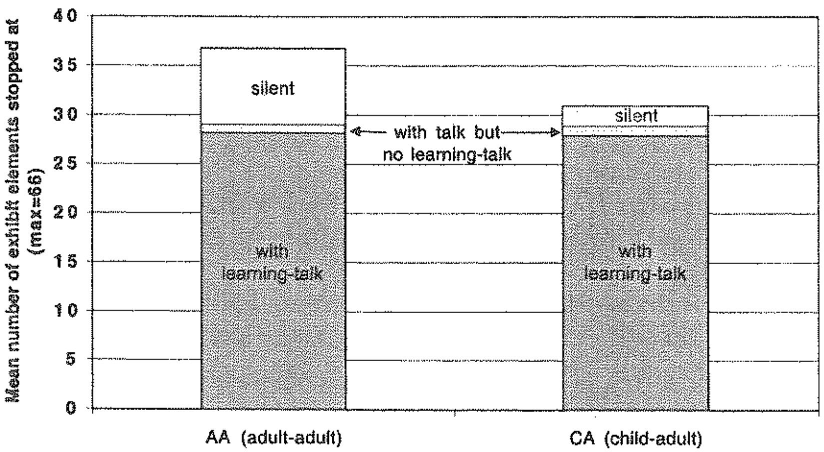

FIG. 8.2. Mean number of exhibit stops by AA and CA dyads.

The AA dyads spent slightly longer than CA dyads (27.1 minutes vs. 23.8 minutes), though this difference is not statistically significant. The AA dyads also saw significantly more elements (37 vs. 31; $t_{27} = 2.52$, $p = 0.02$), as shown by the total column heights in Fig. 8.2.

Figure 8.3 compares the frequency of stopping across the four most common types of exhibit elements in the exhibition. The most attractive exhibit type was the Live Animals; on average, dyads stopped at 74% of the 23 live animals in the exhibition. This was followed by Hands-on and Artifacts (not significantly different from each other). Far less attractive were the Readable elements; dyads stopped at only 14% of these 18 elements which involved reading or looking, but no live or hands-on component.[6]

Overall Frequency of Learning-Talk

On average, dyads engaged in learning-talk at 83% of the exhibit elements at which either person stopped (28 out of the 34). This figure is impres-

[6]The figure does not show stopping frequencies for any of the exhibit types which were only represented by two or three physical examples. These types included Videos (multiple clips of frogs eating or moving), Cases of Organic Material (such as frog foods and dissected or deformed frogs), and Labs (rooms containing many live animals resting or in breeding chambers). These have been omitted from the analysis of exhibit types because each type presents such a small number of opportunities for visitors to stop at them, that we consider the stopping frequencies to be unreliable. However, we can get an idea of their relative attractiveness from their ranking as individual elements. The Videos were intermediate in attractiveness, ranking 24, 33, and 46 out of 66 elements. The Cases of Organic Material were fairly high, ranking 11, 13, and 27. The Lab windows ranked 4 and 14, which made them comparable with the most attractive type of exhibit elements, the Live Animals in smaller terrariums.

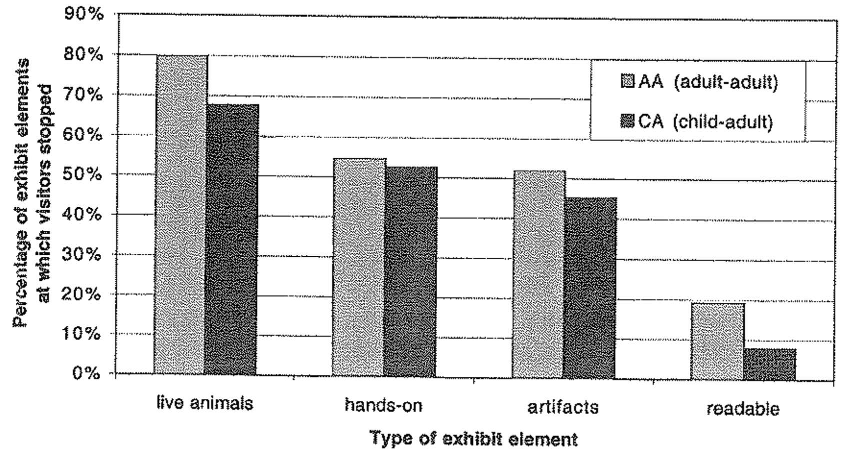

FIG. 8.3. Types of exhibit elements where visitors stopped.

sively high, considering that the visitors in a dyad were free to split up and view exhibit elements alone, if they chose. They had only been instructed to "do whatever you'd normally do," and indeed, many of them did split up for periods during their visit. The high frequencies of learning-talk for both AA and CA dyads (shown in Fig. 8.2) indicate that a broad range of visitors repeatedly chose to engage in this kind of communication with their companion, even in a noisy and stimulating environment where conversation requires some effort.

At the remaining 17% of stops, then, no learning-talk occurred. In the majority of these cases (14% of stops) there was no talk at all; the stop was a silent one. Usually, this reflected a situation where the dyad under study had temporarily split up to pursue individual interests in moving about the exhibition. More importantly, it was quite rare (occurring at the remaining 3% of stops) for a dyad to talk while at an exhibit element but say nothing that could be coded as learning-talk. A closer look at the nature of these "learning-free" conversations reveals that most were expressions of a navigational nature, either beckoning someone to come over, or noting some area of the exhibition that had been missed. Only one case by one dyad could be considered "off-task" in terms of attention to the exhibition. In this single case, a woman said to her adult companion, "I think I hear the girls," which was shortly followed by their exit from the exhibition.

Figure 8.2 compares the occurrence of learning-talk by AA and CA dyads. Learning-talk by AA and CA dyads occurred at almost exactly the same absolute number of stops (28.2 vs. 27.9). Also very similar were the number of stops where there was talk, but no learning-talk (0.9 vs. 1.0). However, AA dyads had significantly more silent stops (7.7 vs. 2.0; $t_{27} = 2.14$,

$p = 0.04$), which may account for the slightly longer times of adult dyads in the exhibition overall. It seemed, then, that AA and CA dyads engaged in learning-talk to a similar extent, but that the AA dyads took a little more time to browse some additional exhibit elements, mostly silently.

Categories of Learning-Talk

Figures 8.4 and 8.5 show the frequencies of the five major categories of learning-talk: perceptual, affective, conceptual, connecting, and strategic. For each dyad, the frequency of each category of talk has been calculated as a percentage of the total number of exhibit elements where either visitor stopped. These percentages have then been averaged across all dyads (Fig. 8.4) or across AA and CA dyads separately (Fig. 8.5).

Overall, the most common categories of talk are perceptual, affective, and conceptual. Pairwise comparisons of the frequencies in each category (summing over all dyads) show significant differences between all but the affective and conceptual categories.

Particularly interesting is the high frequency of conceptual talk. This is heartening news: Museum professionals are constantly looking for evidence that learning is occurring during regular museum visits, yet such evidence is often hard to find if learning is assessed with outcome-based tools imported from the school environment. In this study, we found that visitors engaged in some type of learning-talk at 83% of the exhibit elements they stopped at, and in specifically conceptual talk at 56%.

Also noteworthy is the lower frequency of connecting talk, which includes connections among exhibit elements, connections to previous

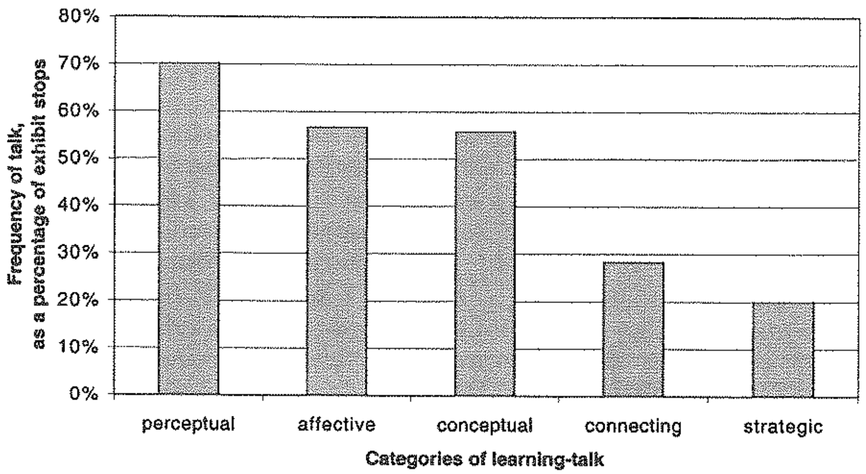

FIG. 8.4. Frequencies of different categories of learning-talk (30 dyads).

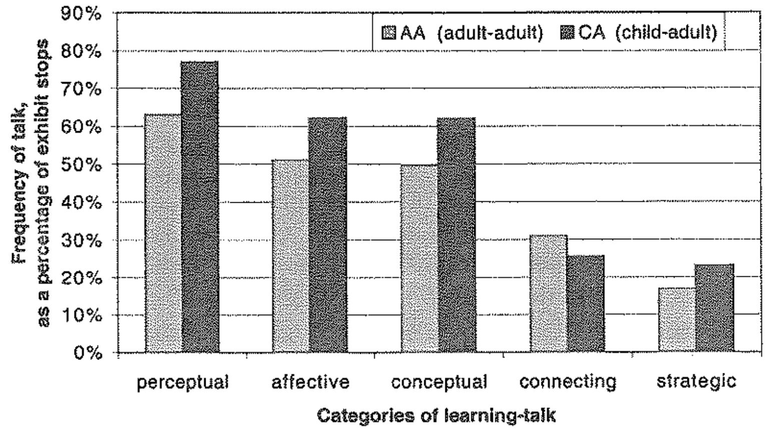

FIG. 8.5. Frequencies of categories of learning-talk for AA and CA dyads.

knowledge, and personal stories or associations. These kind of connections are often regarded as a powerful and ubiquitous means of learning in informal settings, so it is interesting that the category was much less frequent than the "Big 3" (perceptual, affective, conceptual), and was present at only 28% of the exhibit stops overall.

Finally, Fig. 8.5 shows that CA dyads engaged in learning-talk slightly more frequently than AA dyads in four of the five categories. However, this difference was spread almost uniformly across the different categories, and probably reflects the general tendency of AA dyads to view more elements alone, thus lowering their percentage of talk per exhibit stops. The exception is the category of connecting talk, the only one in which AA dyads engaged more frequently than CA dyads. Although the difference is not significant ($t_{27} = 1.17$, $p = 0.25$), the deviation from the pattern in other categories suggests a true (if slight) tendency for adults to make more connections to their prior experience and knowledge than child–adult dyads. It is interesting that this effect was limited to connecting talk, and did not apply to conceptual talk. Perhaps parents felt that the stimulating environment of the exhibition supported immediate inferencing rather than storytelling.

MORE DETAILED ANALYSES

In order to explore the learning-talk data in more depth, I conducted three types of more detailed analysis. First, I divided the five categories of learning-talk into their 16 subcategories, and analyzed their relative frequencies over the whole exhibition visit. Second, I used the main categories to compare the impact of two different types of exhibit element within the exhibition. Third, I ranked the 66 individual exhibit elements according to frequencies of learning-talk, to identify some outstanding individual examples and highlight some of their design features.

1. Subcategories of Learning-Talk

Figure 8.6 shows the frequencies of the 16 subcategories of learning-talk, averaged over all dyads. As in Figs. 8.4 and 8.5, the percentages are out of the total number of elements where visitors stopped, rather than the total number of elements in the exhibition, to remove the effects of differential attracting power. The figure shows a detailed analysis of the different kinds of utterances that contribute to the larger categories of conversation. In this graph, all pairwise comparisons among subcategories give significant differences except for those between identification and quotation, intrigue and pleasure, and simple and complex inferences.

The most common subcategories of learning talk were: identification (44%), quotation (43%), complex inferencing (37%), intrigue (37%), pleasure (36%), simple inferencing (36%), and feature (35%). By contrast, predictions (3%) and inter-exhibit connections (5%) were the least common, and much rarer than we would have expected for a thematic exhibition in an inquiry-based learning institution. These findings are discussed further in the Discussion section later in the chapter.

Voiced Intentions. One of our research questions was: Is there any evidence that the exhibition might affect visitors' attitudes or behaviors once they return to the world beyond the museum? Of course, this is not an easy

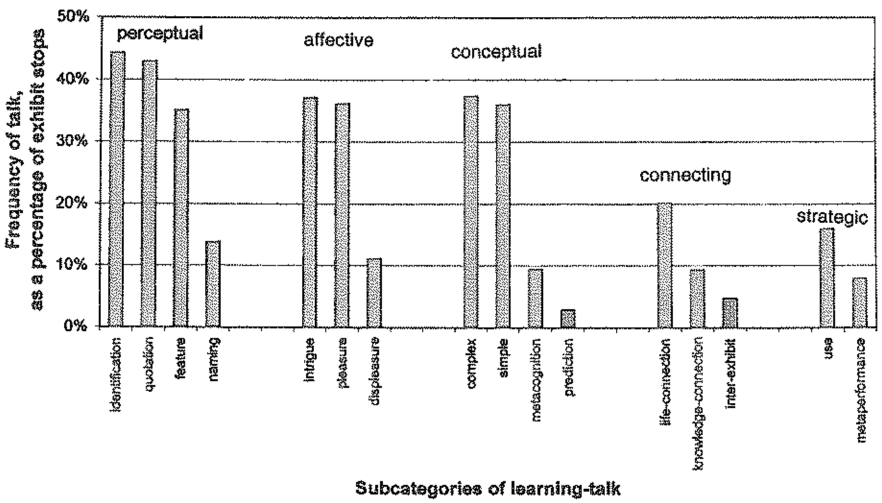

FIG. 8.6. Frequencies of different subcategories of learning-talk.

question to answer, but we were able to code conversations for visitors at least expressing their *intent* to take some action beyond the walls of the exhibition.

In all, 7 different dyads expressed an intention to carry out a total of 10 actions. Of the 10 actions visitors spoke about:

- 3 were intentions to share something from the exhibition with others in some way:

  "I'm going to tell Alice there's a Surinam Toad here."
  "We should take our drawings home."

- 3 were joking expressions of whimsical or teasing intentions that seemed unlikely to happen in reality:

  "In case we meet a frog on the way home, we'll steer clear of his tongue!"
  "I'll buy a load [of frogs] and keep them in the apartment."

- 4 were actions that seemed seriously intended, inspired by different aspects of the exhibition. They were:

  "I should go to Tompkins County when I'm in New York."
  [at an immersion exhibit element where visitors sat on a 'porch in Tompkins County' and listened to frog sounds]

  "Don't eat frog legs any more."
  [at a label about the different animals that eat frogs]

  "You need to keep your frogs at home near water, to keep them from dehydrating."
  [at an element showing the thinness of frog skin and the threat of dehydration]

  "We should get that."
  [to a fellow scientist, after watching a time-lapse sequence of frog embryos]

Although such instances were very rare, it is at least encouraging that they were aligned with the curatorial goals of inspiring visitors to appreciate and protect frogs.

2. Comparison of Exhibit Types

One of the research questions driving our study of the *Frogs* exhibition was: Is there any difference between the patterns of learning-talk that happen at different types of exhibit elements? We thought *Frogs* would provide an excellent opportunity to study this, because the Exploratorium had taken the unusual step of designing an exhibition that combined elements from different museum traditions.

Our original research plan was to compare the patterns of learning-talk at as many of these exhibit types as possible. However, when we began to code and analyze the data, we became aware that our ability to do this would be severely curtailed, due to insufficient numbers of visitor stops at most of the exhibit types. Because our coding scheme allocated a score of 0 or 1 to each exhibit element, we required numerous examples of any given exhibit type in order to generate a distribution of scores by different dyads. The exhibition itself had four types of exhibit element represented by five or more examples: live animals, hands-on exhibits, cultural artifacts, and readable elements. However, two of these types (viz. cultural artifacts and readable elements) were so underused by visitors that the average number of stops at each of these was only 2.4 by each dyad. This made any comparative analysis of little use, because visitors had effectively reduced their own exposure to these types to a point where the frequencies of different types of learning-talk were unreliable.

These limitations were not entirely predictable ahead of time, as they depended in part on the specific choices visitors would make about where to stop in the exhibition. Counting frequencies in a free-choice environment is inevitably susceptible to such disappointments.

We did, however, have two exhibit types (live animals and hands-on elements) that had high enough stopping frequencies to warrant comparison of the learning-talk they elicited. On average, dyads stopped at 17.0 live animals and 5.4 hands-on exhibits.

Averaging over all dyads, visitors engaged in learning-talk at 88% of the live animal exhibits they stopped at, compared with 75% of the hands-on exhibits they stopped at. This difference is significant under a paired t test ($t_{29} = 3.99$, $p < 0.001$), and suggests that even in a hands-on museum such as the Exploratorium, exhibits do not necessarily have to have a manipulative quality in order to evoke widespread learning-talk among visitors. In this exhibition, "hands-*off*" exhibits, in which physical interaction was limited to the search and observation of live animals, elicited slightly higher rates of learning-talk among both AA and CA dyads.

Figure 8.7 shows a more detailed analysis of the differences between learning-talk at live animals and hands-on exhibit elements. All pairwise differences between the two exhibit types were significant. The graph shows that strategic talk was the only category in which visitors engaged more frequently at hands-on elements than live animals. This difference is primarily due to a difference in the subcategory of "metaperformance," in which visitors reflect on their own skill or accomplishment.[7] In every other category,

[7]Strategic talk also includes utterances about how to use an exhibit. Although it might seem almost inconceivable to have a "use" category for live animals, the coding scheme for strategic talk included any reference to *how* to look at something, where to stand, how to locate it, etc. There were no significant differences between exhibit types on the "use" subcategory.

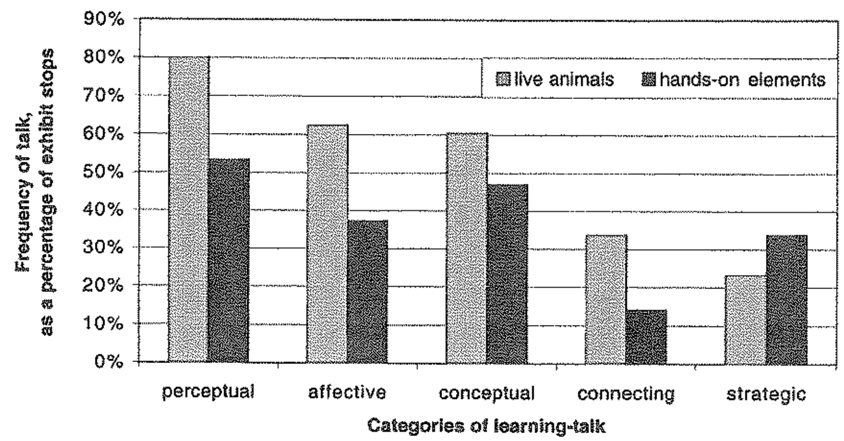

FIG. 8.7. Frequencies of categories of learning-talk at two exhibit types.

live animals evoked more frequent learning-talk than hands-on exhibit elements. In the perceptual category, live animals evoked significantly more talk in the following subcategories: identification, naming, and noting features. In the affective category, live animals evoked significantly more positive affect, and more surprise/intrigue. In the conceptual category, they evoked more simple inferences, complex inferences, and metacognition. And in the connecting category, live animals evoked more life connection and knowledge connection. These findings show that, at least in this exhibition, live animals not only stimulated more frequent learning-talk overall than hands-on elements, but that the differences were much more widespread than purely perceptual modes of learning.

3. Outstanding Individual Exhibit Elements

In addition to providing overall patterns in the conversation data, the coding scheme can be used to identify individual exhibit elements which were outstanding in terms of their tendency to evoke certain subcategories of learning-talk. We highlight some of them here to give texture to the data, and also to highlight key design characteristics that we think may have led to the success of each one as elicitors of learning-talk. In each case, we have selected the exhibit element (out of 66) that gave the highest frequency of talk by all the dyads who stopped there.

Quotation. The most-quoted label was that of "Frog's Eggs," an element designed to show the embryonic development of fertilized frogs' eggs up to the time of hatching. This exhibit element displayed a series of jars containing real frog embryos at different stages of development, as well as a timelapse video that condensed two days of development into a continuous sequence lasting one minute.

Examples:

"Look through the magnifier to see frog's eggs in several stages of development."

"Three days, and at the end it's five days . . . these are frog eggs."

"That's 21 hours, 27 hours . . . 2 days, 5 days . . ."

"Frog's eggs . . . Look at the African frogs mate in the fall. Hundreds of eggs are laid and then it only take them 3 or 4 days to hatch . . . they use time-lapse photography."

"In 3 months."

We hypothesize that this element was particularly likely to elicit quotations because it contained luminous displays of compelling organic material without narration, so that visitors who were initially drawn to the visual stimuli wanted a reference or explanation of what they were seeing. Also, the series of jars invited comparison by visitors, but the dimensions of comparison were fairly technical in nature (e.g., "21 hours: developing backbone visible"), thus leading to a tendency to quote verbatim from the label. Finally, the fact that this was one of the earliest elements in the exhibition may have made visitors more likely to read the label aloud; museum researchers (e.g., Falk, Koran, Dierking, & Dreblow, 1985; Gilman, 1916) have noted the robust phenomenon of "museum fatigue," in which visitors view exhibits more cursorily after a period of time, and this effect may apply to an exhibition as well as a whole visit.

Complex Inference. The element evoking most frequent complex inferencing was "Frogs and Toads." This was a large open tank containing a miniature ecosystem, including a small stream, stony beach, and dense vegetation. It also contained 17 marine toads that, being nocturnal, were notoriously difficult to locate during the day. Typically, visitors searched for the toads for some time, and then generated hypotheses for why the toads might not be visible. In our coding scheme, all hypotheses were regarded as complex inferences.

Examples:

V1: Would they bury themselves?

V2: Perhaps, yeah, or they may be really camouflaged too.

"Maybe it's just showing where they like to live."

"Something must be under here because, see, the water is moving."

This element was also very early in the exhibition (usually the first live exhibit visitors encountered), and this may be partly responsible for its success in eliciting inferences. However, it is striking that the most frequent complex inferencing was at an exhibit where the phenomenon described in the label was often impossible to see; it seems that the lack of obvious frogs served both to slow visitors down, and to generate a mystery or challenge worthy of discussion. In this sense, the element facilitated learning in an unexpected but fruitful manner.

Prediction. A silent, 90-second video called "Mealtime" showed a series of clips of frogs catching and eating a variety of foods, including insects, maggots, and even a live mouse. Most visitors who stopped at the video stayed for a complete cycle, and it was not unusual for them to stay until a favorite clip came around again.

Examples:

"Oh, they're going to start it again."

"See, they're gonna catch their food."

"Beetle, bet you it spits it out."

The repeating nature of the video seemed to encourage visitors to make predictions, especially when only one member of the dyad had seen the footage the first time. Also, the label invited them to "see what happens when a frog accidentally eats something nasty," implicitly inviting them to make a prediction.

Metacognition. The Mealtime video was also the element most likely to evoke a metacognitive comment from visitors.

Examples:

"Well, I would be surprised that it ate a whole mouse."

"I never would have believed."

"I didn't realize they got them with their tongue. Did you?"

I believe the success of this element was due to its surprising, articulable content message, and its strong graphic presentation. Visitors were not aware of the size or variety of foods that frogs eat, and the nature of the video footage (colorful, high-speed and suspenseful) made this point so powerfully that visitors were eager to share their surprise with each other.

Life Connection. The most frequent personal connections were elicited by a graphic reproduction of a leaf from the children's book, *Frog and Toad are Friends.*

Examples:

"Oh, this is, oh, look, it's Frog and Toad. I remember that one. That was your favorite book when you were little."

V1: 'Cos here's the story, Mom. I used to have these books a lot.
V2: Oh you're right. Oh, you're exactly right. That's Toad and Frog.

The success of this particular element reflects the use of evaluation studies to inform the choice of popular book. By choosing a well-known children's book, the development team hoped for exactly this response, that visitors would be reminded of their own personal connection to frogs and froggy stories.

Knowledge Connection. The element most likely to generate connections to previous knowledge was a bulletin board of current events called "Frogs and Toads in the News." Most of the articles on the board concerned the scientific debate about whether declining frog populations were caused by human activity such as the use of carcinogenic chemicals. Although few visitors stopped there, most were highly informed about frogs, so this element had the highest ratio of knowledge-connection talk per stop.

Examples:

V1: I thought they found that they don't believe there's any type of carcinogen or anything that, because it's . . .
V2: Well, it depends on where it is though.
V1: it occurs so naturally and so frequently.
V2: Oh, I see what you're saying. That I don't know.

V3: So it splits the developmental field at a time when it can regenerate PTA.
V4: I guess at some point it prevents two groups of cells from talking to each other while one is, they are supposed to form one thing, they form three things. Most split . . .
V3: Because they form two out of the same, but, that . . .
V4: Yes. There's some little undergrad who did it at Stanford.

It is interesting that a bulletin board displaying current science articles should elicit more of visitors' own declarative knowledge ($n = 4$) than quotes from the articles ($n = 0$).

Inter-exhibit Connection. The only element that evoked more than three inter-exhibit connections was the live "Poison-Arrow Frog." In all, the *Frogs* exhibition contained three terrariums of live poison-arrow frogs, which

were not all the same species, and had a variety of colorations. Nine visitor dyads made explicit or implicit reference to the fact that they had already seen a poison-arrow frog in a previous terrarium.

Examples:

V1: Oh Mum, look at this frog.
V2: Oh, poison-arrow. These are my favorite ones.

"Those are more poison-arrow frogs."
"That's a poison-arrow too."

Clearly, visitors made this kind of connection because different frogs had the same common name on their label. While being a tribute to careful label reading by visitors, this type of inter-exhibit connection was probably more superficial than the development team would hope for.

Use. The sound exhibit "Croak like a Frog" generated the highest frequency of utterances about how to use an exhibit. In this exhibit, visitors could listen to a variety of prerecorded frog calls, and record their own attempts to imitate these calls.

Examples:

"You gotta record your own, honey. Now, you push the record button."
"You have to do it before the red line disappears, or it doesn't record."

I believe this exhibit elicited most use-talk because it had a combination of a clear, appealing goal, and a fairly complex interface. The title, "Croak like a Frog," along with the installed microphone, probably served to give visitors easy access to the main point of the activity, and engaged them enough to help each other through the subtleties of the push-button interface.

Metaperformance. "Croak like a Frog" also generated the highest frequency of reflection on skills or performance.

Examples:

V1: What do you think, Jill, close or not close?
V2: Not close.

"That's not too bad, try this one."
"See, look at the difference . . . It's different. Okay! This was right, except I made it too long."

As well as playing the sound of a frog or human imitator, this exhibit included a graphical display of the volume, frequency, and timing of the sound on a computer screen. For this reason, it afforded (and explicitly encouraged) both visual and aural comparison of the authentic frog song with one's own attempt. In other words, metaperformance was a central design goal of this element, and visitors' conversations verified its success in this regard.

Pleasure. The exhibit that generated the most frequent expressions of pleasure was "African Clawed Frog," a spectacular large tank of swimming *Xenopus laevis.* Visitors' laughter accounted for many of the coded examples.

Examples:

"They're funny."
"That's hilarious. I want to hug you . . ."
"Cool!"
"I like this one. Yeah, that is so cute."

I would speculate that the main reason for visitors' widespread pleasure with this element was due to its unusually lively inhabitants. For most of the exhibition, terrariums contained frogs at rest, but this huge tank contained frogs that were alternately perfectly still and actively swimming. In addition, the swimming frogs displayed a particular behavior that many visitors noticed and found endearing: they stretched one arm out in front of them as they swam, as if to warn them of unexpected obstacles.

Displeasure. The exhibit that elicited the most frequent expressions of displeasure was a plastinated dead frog displayed to show the internal organs. It was labeled "What's Inside a Frog?"

Examples:

"Eeww! Oh my gosh, okay. Eeww."
"Yuck, where's his brain?"
"Don't care to see the insides."
"Poor frog. Eew."

Perhaps this negative affect was the most predictable of responses, because the exhibit element combined graphic organic material (brightly colored organs) with the context from which it came (the surrounding frog, roughly intact). Thus we recorded visitors who found the exhibit either gory or sad, or both. The development team had previously argued about whether to include dissected frogs in the exhibition, given its focus on respect for these beautiful living animals. The final decision was made to include a plastinated frog for the scientific value, and to make the point that frogs and humans have so much in common.

Overall Diversity. By counting the number of codes assigned to individual exhibit elements, it was possible to determine which element elicited the largest number of different subcategories of learning talk, on average. The video element called Mealtime evoked an average of 6.1 different subcategories of learning-talk. This seems a particularly impressive total, given that the exhibit is neither live nor hands-on, nor does it even have an audio component. This finding shows that it is possible to use video to elicit a broad range of learning conversations among visitors.

Although a detailed analysis of rankings is beyond the scope of this chapter, we note that, after the Mealtime video, the following 14 top-ranking elements in terms of diversity all included live animals (either in individual terrariums or in the two laboratory windows). The hands-on elements were intermediate in the diversity of learning-talk they elicited, and the readable elements (including the exhibition introduction, maps, bulletin boards, and highlights of books and stories about frogs from a variety of cultures) ranked extremely low. All 18 readable elements fell within the lowest 25 rankings out of the 66 elements in the exhibition, in terms of diversity of learning-talk by those visitors who chose to stop there.

DISCUSSION

Similarities to Findings From Previous Studies

Frequency of Learning-Talk. Overall, learning-talk in *Frogs* was extremely frequent, occurring at 83% of the elements where either member of a dyad stopped, and fully 97% of the elements where some kind of talk occurred. Comparable figures were obtained by Hilke (1989), who reported that 86% of all events undertaken by families concerned the exhibits, and that 64% of all behaviors (including nonverbal behaviors) were exhibit-related behaviors that would classify as learning-related in the current study. Our data supports Hilke's general conclusion that families are "pursuing a clear agenda to learn while in the museum." Even more strikingly, we found an extremely low percentage (3%) of exhibit stops where visitors discussed something that did not contribute to exhibition-related learning. Although it must be acknowledged that the effect of cueing visitors may have had an effect here, I believe it unlikely that a dyad could stay "on-task" to this degree for nearly half an hour unless it was close to their natural behavior. Our data support McManus's (1989a) conclusion that "Conversations were about the topic established by label-writers. Instances where visitors shifted from the exhibit topic to introduce an unrelated topic were exceedingly rare in the transcripts" (p. 181).

Frequency of Quotations From Labels. We found that label-reading was common: On average, visitors quoted from labels (or closely paraphrased them) at 43% of their exhibit stops. Although many of these cases were simply a few words, it shows that visitors were attending to labels to a high degree. McManus (1989a), whose study dispelled the common notion that "visitors don't read labels," found an even higher rate among four exhibit elements at the British Museum (Natural History): an average of 71%. Diamond (1980), on the other hand, reported an average closer to 10% of exhibit elements stopped at, but she used a narrower definition of "read aloud" that excluded paraphrasing, so we would expect her frequencies to be considerably lower than those in our study.

Comparability With Exhibitions From Art and History Museums. Although developed independently, our coding scheme turned out to be surprisingly compatible with Silverman's (1990) earlier analysis of visitors' conversations in an art and a history museum. Silverman used very similar methods to those we used to gather conversational data from dyads of visitors, and proposed five broad categories of interpretive speech acts: establishment, absolute object description, evaluation, relating special knowledge, and relating personal experience. Her categories correspond closely to our categories of: perceptual talk (excluding the "feature" subcategory), feature talk, affective talk, knowledge connection, and life connection.

Disparities and Unexpected Findings

High Frequency of Conceptual Talk. In our greatest divergence from Silverman's coding scheme, we identified Conceptual Talk as a major category of learning-talk, and one of the three most common in visitors' conversations. This category included the subcategories of prediction, metacognition, simple, and complex conceptual talk. The most common subcategory was "complex inferencing" involving hypotheses, generalizations, or relational thinking. This occurred at 37% of exhibit elements where either visitor stopped, and shows that even inferential reasoning, though often brief and informal, is not uncommon in visitors' conversations.

Silverman's coding scheme does include two categories where conceptual talk could be counted.[8] However, in both cases the inferences and hypotheses are treated as minority members of a larger category with a different focus. This divergence in coding schemes may reflect subtle differences in the researchers' epistemologies and areas of interest. Alternatively, it may reflect a true difference in visitors' conversations in the different types of museums represented. Perhaps visitors to a hands-on science museum are more likely to engage in verbal acts of reasoning, prediction, and self-reflection than visitors to either an art or history museum, by virtue of the different missions and natures of these institutions.

[8]The categories are: a) "interpretation," a subset of "evaluation," where visitors share a message, meaning or conclusion derived from an object; and b) "absolute object description" which includes not only perceptual aspects of an object, but elaborations and deductions about their function and subject matter.

Perhaps Hilke (1989) comes closest to including a conceptual category, which she names "pure-info" and which contains "action-events that involve direct attempts at acquiring or exchanging information" (p. 110). However, Hilke's subcategories for this code include nonverbal behaviors such as "gaze at" and "manipulate," which, while evidence for engagement with the exhibit, would fall outside our narrower definition of verbal evidence for learning.

I would argue that learning-talk of a conceptual nature should be highlighted in museum research, especially because it is the category that comes closest to traditional definitions of learning from school and laboratory settings. While museums may have unusual strengths in facilitating learning of the affective or sensory kinds, it is also important to recognize that visitors engage frequently in cognitive learning-talk during their exhibition visits.

Strategic Talk. We also identified the category of Strategic Talk, incorporating talk about how to use an exhibit element, and reflections on one's degree of success in using it.

Several other researchers have coded this kind of talk somewhere in their scheme, but have coded it together with all kinds of comments about an exhibit (e.g., Borun et al., 1998) or with metacognitive statements such as "I understand it" (e.g., Diamond, 1980). Hilke (1989) included the category of "Say what to do," which may have coded talk about how to use an exhibit, or may also have included beckonings and turn-taking kinds of statements, which we did not code as part of "use-talk." Interestingly, Silverman (1990) did not mention any strategic category or subcategory. Perhaps this reflects a real difference between the kinds of learning that happen in a hands-on versus more observational museum exhibition; part of what is learned in a hands-on exhibition is how to interact with the objects it contains, with skill and self-awareness.

I believe that strategic knowledge, about how to interact with exhibition objects, is a category of learning-talk worth recognizing. In attempting to extend its applicability beyond manipulable exhibit elements, we included in this category any utterances related to *how* to see, hear, or otherwise interact with, any exhibit element, even the live frogs in terrariums that needed skill to locate. For this reason, I propose it as a generalizable category of learning-talk.

Rarity of Connections Among Exhibits. Visitors made explicit connections to other exhibit elements at only 5% of the elements (1.6 elements on average). This result is surprising and quite provocative, given the large amount of focused effort that goes into the design and coordination of objects, labels, graphic treatments, and visitor orientation within a typical museum exhibition.

It is interesting that Diamond (1980) did not have "inter-exhibit connections" in her category system, even though she did have "reminisce about exhibit," in which a visitor makes a comment about a previous interaction with the same exhibit. Perhaps it is rarer for visitors to make connections among exhibits than it is for them to make connections among different interactions with the same exhibit.

Hilke (1988) interpreted her own data in a way that may shed light on ours: "The family's primary agenda is not to look for relationships within the content of the show. Rather, family visitors will seek relationships between their own knowledge/experience and the content/structure of the show. The dominant perspective from which the exhibition is interpreted is more likely to be the visitor's own background experience, own knowledge, and own interests than it is likely to be some common thread or theme of the show" (p. 124). Taylor (1986) reached similar conclusions.

It may be argued that visitors did not explicitly state all of the connections they were making among elements, but this argument could be applied to other types of learning-talk as well, and would not explain why inter-exhibit connections were the second-rarest of the coded subcategories. It is true that this subcategory was coded conservatively, meaning that the coder had to feel confident that the visitor was making a connection to an exhibit element rather than to more general previous knowledge. It may even be the case that the design of the *Frogs* exhibition was flawed, making it unusually difficult for visitors to recognize links among its elements. Serrell (1996) made a strong case for having exhibition design be driven by a single "big idea" that should not be vague or trivial. She wrote, "The big idea provides an unambiguous focus for the exhibit team throughout the exhibit development process by clearly stating in one non-compound sentence the scope and purpose of an exhibition" (p. 2). From this standpoint, *Frogs* could perhaps be criticized as having had multiple goals that were difficult to combine into a single coherent idea; the exhibition was much more of a potpourri of interesting aspects of frogs and their relationship to people.

However, even with all these caveats, the frequency of this type of talk is so low compared with other types that I believe it warrants further attention in the research community.

Rarity of Predictions. Visitors made explicit predictions at only 3% of the exhibit elements (an average of 1.0 element). This was the rarest of the subcategories coded, and seems unexpectedly low, especially in a museum that

emphasizes inquiry processes and engagement through curiosity and surprise (e.g., DeLacôte, 1998; Oppenheimer, 1986; Semper, 1990).

It may be that predictions were especially low in the *Frogs* exhibition because the live frogs were relatively static, and did not support direct intervention (except tapping on the glass, which was discouraged in signage and by museum staff).

Alternatively, it may be an unavoidable artifact of our methods: our microphones can record speech but not thought. Perhaps museum visitors are predicting at a more kinesthetic or experiential level, bringing unconscious expectations of the way the world works to their evolving interaction with an exhibit. In the fast-paced, social, stimulating environment of a hands-on museum, predictions may simply take too much time or be too effortful to fully articulate. Perhaps visitors' expressions of surprise and intrigue ("Wow!") should serve as better verbal indicators that an unspoken prediction has been disconfirmed. More light could be shed on this issue if we had comparable data from more typical hands-on exhibits.

Comparison of Types of Exhibit Element: Live Versus Hands-On. Our study showed that live animal exhibit elements elicited significantly more frequent and also more diverse learning-talk than hands-on elements. This is another surprising and provocative result, given that the educational lore of museums, backed by an array of studies, gives such high value to interactive experiences. How is it that small, generally inactive animals confined to terrariums in which they could not be touched or communicated with, could elicit more learning-talk in four out of five categories than custom-designed, hands-on exhibits?

I believe the data constitute a challenge to simplistic notions of interactive learning. Several museum researchers have argued against too narrow an interpretation of concepts such as "interactive" or "constructivist." Hein (1998), while listing some of the research on the value of adding manipulative components to museum exhibits, made the strong point that learning is increased by "*meaningful* physical activity." Osborne (1998) pointed out that "experience, of itself, while highly enjoyable, is overwhelmingly a missed learning opportunity without some attempt to encourage the visitor to focus, recapitulate, and review" (p. 9). As McLean (1993) put it, "rows of buttons and levers may exercise some visitors' fingers and arms, but not necessarily their minds" (p. 16).

With the terrariums of living frogs, the developers seem to have achieved the opposite situation: exhibit elements that are "minds-on" but "hands-off." Although almost entirely noninteractive, the living frogs attracted more visitors and inspired more learning-talk than their interactive neighbors. Of course, even without manipulation, the live animal displays still offered visitors a powerful direct experience on which to reflect, an experience quite different from reading a book. The terrariums were designed to look like jewels, full of rich color and texture from the frogs, foliage, pebbles, sand, and curving tree branches. The visual beauty and complexity of these tiny worlds invited visitors to point things out to each other, and to share their surprise and delight on finding a brilliant Golden Frog or craggy Surinam Toad. The hidden quality of many of the small frogs invited visitors to slow down, look carefully, and hypothesize about what they were seeing or not seeing. The labels, after some formative evaluation, answered some of the questions that often arose in visitors' minds. This was fertile ground for learning-talk.

Several of the hands-on elements had an audio rather than a visual focus. For example, visitors could listen to songs about frogs, hear frog calls, or try their hands at imitating frogs. Perhaps the audio experience tended to inhibit visitors' conversations at these elements (audio was noticeably absent from the highly successful Mealtime video). Visitors wearing headphones, even with the same sound piped to each, would have had to put more effort into sharing any kind of response, thought, or feeling. Also common in the exhibition were hands-on elements that involved interacting with symbolic objects that represented frogs in some way. For example: A zoetrope showed the detailed action of frog locomotion; a set of maps could be lifted to show the changing populations of frogs with time; or a set of door-panels could be opened to reveal the inner anatomy of a frog as compared with a human. Perhaps these symbolic frogs were less compelling, less inspirational, and less evocative of associations, than the "real thing."

Frogs was just one exhibition, and the findings may not generalize even to other Exploratorium exhibitions. Nevertheless, this case does show that designing for learning conversations does not depend on a simple variable such as whether or not exhibit elements are hands-on. I hope museum researchers will continue to study the subtleties of effective exhibit design in the exhibition context.

ANALYZING VISITORS' CONVERSATIONS AS A METHOD FOR VISITOR STUDIES

To end this chapter, I offer some reflections on the advantages and drawbacks of the methods used to studying visitors' learning through the lens of their real-time conversations while in an exhibition.

Hearing or reading visitors' complete conversations is a vivid experience that brings one right into the arena where real museum learning occurs. The transcripts are detailed, dense, and at times brutally honest, providing readers (be they developers, evaluators, or researchers) with a gritty sense of what engages and what doesn't. Personally, I found it a striking reminder

of the power of choice in informal environments: Visitors are choosing where to spend every second of their time, and exhibits that do not engage or sustain them are quickly left behind, however "potentially educational" they may be.

Following a single group through an entire exhibition provides continuity; over that length of time, one begins to get a sense of the visitors' particular personalities and preferences, as well as their ways of interacting with each other. Falk and Dierking (1992) called this the "personal context" and "social context" in which museum visits occur, and which interact with the "physical context" to create the visitor experience. It is much easier to understand visitors' personal or social contexts when studying a half-hour of their conversation than the few minutes typical of a single-element interaction.

For the educational researcher, visitors' conversations provide raw material for multiple kinds of studies: comparisons between types of exhibits or exhibitions to explore the effects of different learning environments; studies of sequential patterns in the talk; typing or characterization of visitors based on their talk; and search for possible correlations among conversational talk, visitors' answers to structured interview questions, time spent, and physical interactivity.

In choosing an operational definition of learning-talk, a researcher can choose one of two main directions. One can either define learning with respect to visitors' previous knowledge, or with respect to the contents of the exhibition. The former is difficult because of the practical and theoretical challenges of pre- and posttesting visitors on a vast range of possible ideas and associations which they may bring to bear on their visit. The latter is difficult because it requires that coders be intimately familiar with every word and image in an exhibition, in order to determine the nature of the learning-talk which has just occurred. Ironically, this becomes more difficult as exhibit elements become more effective, because when visitors' language and label language mesh seamlessly, it is hard to tell when visitors are quoting.

For the museum evaluator, visitors' conversations provide plenty of detailed summative information on visitors' responses to individual exhibit elements. The original recordings or transcripts provide a wealth of information for anyone on the exhibition team willing to spend the time to get a close "feel" for the different visitors' experiences. With the addition of a systematic analysis into categories, conversation studies can provide comparisons among the impacts of different types of elements. These studies can also identify the most successful elements in an exhibition, based on criteria such as pleasure, intrigue, and life connection, which are often explicit goals of exhibit development teams. Seeing the "winners" in each category is useful both as a reality check for discussions of exhibit design features (even in the absence of a controlled design experiment), and as data for developers who develop intuitions about visitor behavior mostly through induction over many examples and many years of experience. Even more important, these studies bring the focus of the design debates to the real-time visitor experience, and out of the domain of theory, rhetoric, or even simple observation. In doing so, they challenge developers' assumptions about what visitors will find meaningful in their exhibitions.

One disadvantage of this kind of study is that visitors are behaving inauthentically, at least to some degree, because they are aware of the microphones. We have made an argument for regarding visitors' behavior in this situation as "best case," but if that is so, we need to remember it when interpreting the findings, and resist the temptation to over-congratulate ourselves on our success, particularly in the absence of comparative data from other exhibitions.

The main disadvantage of this kind of study is that it is slow and expensive. In spite of my original intention to find a method that could translate into a "quick and easy" form for summative evaluation, I have come to believe that conversations are too complex to allow for speedy yet meaningful analysis. Even at best, our coders took four hours to code each hour of transcript: two to parse it into episodes based on exhibit element (an easy task), and another two to code into the 16 subcategories of learning-talk (much more difficult). This figure excludes time needed for transcription and data gathering, both of which are also slow. Slowest of all was the development of a coherent coding scheme reliable enough to be of use.

Stubbs (1983) was right to warn of the difficulties of any kind of discourse analysis in a complex setting. We found visitors' language to be subtle, ambiguous, and incomplete. Humor, which was a common feature of visitors' conversations, was particularly difficult to code.[9] The status of questions, interactions with strangers, and nested kinds of learning-talk, created additional challenges to the coding scheme. Even counting was difficult: our choice of a binary scheme by which to count learning-talk at each exhibit element, while it simplified coding, made reliability difficult to quantify and limited the kinds of quantitative comparisons we could make.

How could one simplify this situation and make it more feasible as an evaluation strategy? Reducing sample size is not an option, as 30 dyads were already fewer than most evaluators would consider for a summative evaluation. One could simplify the coding scheme to, say, the main five categories of learning-talk, but the findings would suffer greatly in terms of resolution, and coding time would not be reduced by very much over all, because even these require fine discrimination to use reliably. One could also omit the

[9]For example, one visitor startled a group of tadpoles when he approached, and joked "Swim for your lives!" Another visitor watched a giant water snail and commented, "Serious escargot." Such comments, while understandable and entertaining, were very difficult to code in our scheme.

transcription step, and code directly from the audio recordings, but our experiments with this suggested that it would result in a critical loss of reliability, with the added disadvantage of losing the opportunity to show a written record with textured examples to the development team.

I would argue that, to really benefit from studies of visitor conversation, the field of visitor studies would need to establish a standardized coding scheme, somewhat analogous to the convergence of tracking and timing methods following the comparative analysis of exhibitions by Serrell (1998). This would save the large amount of time required to develop a coding scheme for each exhibition. It would also allow for a much clearer comparison of exhibition data across sites, designs, and times than is currently possible, allowing both researchers and evaluators to learn much more from our initial investments of time and effort. I am hopeful that this may emerge from the Museum Learning Collaborative, which has been exploring a variety of methods and coding schemes to adopt for a larger multisite study. Even with such standardization, however, there would be much work to do to train individual coders to use the scheme reliably.

In short, this study leads me to the view that analyzing real-time visitor conversations in exhibitions is a fertile but costly complement to more traditional methods from visitor studies, such as tracking and interviewing. Its strength is in bringing the researcher into the heart of the learning "action" of the museum visit, and emphasizing learning as process rather than merely outcome. Visitors' raw transcripts are rich and revealing, but transcription is difficult and readers tend to drown in detail or be overinfluenced by anecdotes. On the other hand, creating summaries of learning-talk, averaged over many visitors, requires careful coding and counting, with no obvious short-cuts.

ACKNOWLEDGMENTS

This research was conducted as part of the Museum Learning Collaborative and was funded by the Institute of Museum Services, the National Endowment for the Arts, the National Endowment for the Humanities and the National Science Foundation.

I would like to express my thanks to Marni Goldman, Joshua Gutwill-Wise, and Isabel Halm for valuable assistance with many aspects of the data-gathering and coding, and to an enthusiastic team of volunteer transcribers: Helen Allen, Karen Dodson, Ruth Erznoznik, Christine Keppler, Ralph Kerwin, Kanani Lee, and Rebecca Lee. My appreciation also goes to Joshua Gutwill-Wise, George Hein, Kathleen McLean, Beverly Serrell, and Raycyn Wright for their thoughtful comments on earlier drafts of this chapter.

REFERENCES

Bitgood, S., Serrell, B., & Thompson, D. (1994). The impact of informal education on visitors to museums. In V. Crane, H. Nicholson, M. Chen, & S. Bitgood (Eds.), *Informal science learning* (pp. 61–106). Washington, DC: Research Communications Ltd.

Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives: The classification of educational goals, Handbook 1, The cognitive domain, Handbook 2, The affective domain.* New York: David McKay.

Borun, M., Dritsas, J., Johnson, J. I., Peter, N. E., Wagner, K. F., Fadigan, K., Jangaard, R., Stroup, E., & Wenger, A. (1998). *Family learning in museums: The PISEC perspective.* Philadelphia: Philadelphia/Camden Informal Science Education Collaboration.

Contini, H. (1999). [A tracking study of the Exploratorium's "Frogs" Exhibition]. Unpublished raw data.

Crane, V. (1994). Understanding the dynamics of informal learning. In V. Crane, H. Nicholson, M. Chen, & S. Bitgood (Eds.), *Informal science learning* (pp. 107–191). Washington, DC: Research Communications Ltd.

Crowley, K., & Callanan, M. (1998). Describing and supporting collaborative scientific thinking in parent-child interactions. *Journal of Museum Education, 23*(1), 12–17.

Crowley, K., Galco, J., Jacobs, M., & Russo, S. R. (2000). *Explanatoids, fossils, and family conversations.* Paper presented as part of a paper set, Museum Learning Collaborative: Studies of Learning from Museums, at the annual meeting of the American Educational Research Association, New Orleans.

Delacôte, G. (1998, June). Putting science in the hands of the public. *Science, 280,* 2054–2055.

Diamond, J. (1980). *The ethology of teaching: A perspective from the observations of families in science centers.* Unpublished doctoral dissertation, University of California, Berkeley.

Falk, J. H., Koran, J. J., Dierking, L. D., & Dreblow, L. (1985). Predicting visitor behavior. *Curator, 28,* 249–257.

Falk, J. H., & Dierking, L. D. (1992). *The museum experience.* Washington, DC: Whalesback.

Falk, J. H., & Dierking, L. D. (2000). *Learning from museums: Visitor experiences and the making of meaning.* Walnut Creek, CA: AltaMira Press.

Garvey, C. (1984). *Children's talk.* Cambridge, MA: Harvard University Press.

Gilman, B. I. (1916). Museum fatigue. *The Scientific Monthly, 12,* 62–74.

Guberman, S. R., Emo, K., Simmons, S., Taylor, J., & Sullivan, G. (1999). *Parent-child conversations in a natural history museum.* Paper presented at the annual meeting of the American Educational Research Association, Montreal.

Guberman, S. R., & Van Dusen, A. (2001). *Children's investigations in a science discovery center.* Paper presented at the annual meeting of the American Educational Research Association, Seattle.

Hein, G. E. (1998). *Learning in the museum.* London: Routledge.

Hensel, K. A. (1987). *Families in museums: Interactions and conversations at displays.* Unpublished doctoral dissertation, Columbia University Teachers College.

Hilke, D. D. (1988). Strategies for family learning in museums. In S. Bitgood, J. T. Roper Jr., & A. Benefield (Eds.), *Visitor studies: Theory, research, and practice* (pp. 120–125). Jacksonville: Center for Social Design.

Hilke, D. D. (1989). The family as a learning system: An observational study. *Marriage and Family Review, 13,* 13–34.

Jeffery-Clay, K. R. (1998). Constructivism in museums: How museums create meaningful learning environments. *Journal of Museum Education, 23*(1), 3–7.

Leinhardt, G., & Crowley, K. (1998). Museum learning as conversational elaboration: A proposal to capture, code and analyze museum talk. (Museum Learning Collaborative Technical Report MLC-01). Available at: *museumlearning.com*

Lucas, A. M., McManus, P., & Thomas, G. (1986). Investigating learning from informal sources: Listening to conversations and observing play in science museums. *European Journal of Science Education 8*(4), 341–353.

Martin, L. M. W. (1996). *A Vygotskian approach to the design of a science center.* Paper presented at the annual meeting of the American Educational Research Association, New York City.

Matusov, E., & Rogoff, B. (1995). Evidence of development for people's participation in communities of learners. In J. D. Falk & L. D. Dierking (Eds.), *Public institutions for personal learning: Establishing a research agenda* (pp. 97–104). Washington, DC: American Association of Museums.

McLean, K. (1993). *Planning for people in museum exhibitions.* Washington, DC: Association of Science-Technology Centers.

McManus, P. M. (1987). It's the company you keep . . . the social determination of learning-related behaviour in a science museum. *International Journal of Museum Management and Curatorship, 6,* 263–270.

McManus, P. M. (1988). Good companions: More on the social determination of learning-related behaviour in a science museum. *International Journal of Museum Management and Curatorship, 7,* 37–44.

McManus, P. M. (1989a). Oh, yes, they do: How museum visitors read labels and interact with exhibit texts. *Curator, 32*(3), 174–189.

McManus, P. M. (1989b). What people say and how they think in a science museum. In D. Uzzell (Ed.), *Heritage interpretation* (pp. 174–189). London: Bellhaven Press.

Munley, M. E. (1992). Back to the future: A call for coordinated research programs in museums. *Patterns in practice* (pp. 196–203). Washington, DC: Museum Education Roundtable.

Oppenheimer, F. (1986). *Working prototypes: Exhibit design at the exploratorium.* Exploratorium Publications, ISSN-0-943451-081-06.

Osborne, J. F. (1998). Constructivism in museums: A response. *Journal of Museum Education, 23*(1), 8–9.

Roberts, L. (1990). The elusive qualitities of "affect". In B. Serrell (Ed.), *What research says about learning in science museums* (pp. 19–122). Washington, DC: Association of Science-Technology Centers.

Semper, R. J. (1990, November). Science museums as environments for learning. *Physics Today,* 50–56.

Serrell, B. (1998). *Paying attention: Visitors and museum exhibitions.* Washington, DC: American Association of Museums.

Serrell, B. (1990). *What research says about learning in science museums* (pp. ii, 30). Washington, DC: Association of Science-Technology Centers.

Serrell, B. (1996). *Exhibit labels: An interpretive approach.* Walnut Creek, CA: AltaMira Press.

Silverman, L. H. (1990). *Of us and other "things": The content and functions of museum talk by adult visitor pairs in an art and a history museum.* Unpublished doctoral dissertation, University of Pennsylvania.

Stubbs, M. S. (1983). *Discourse analysis: The socioliguistic analysis of natural language.* Chicago: University of Chicago Press.

Taylor, S. M. (1986). *Understanding processes of informal education: A naturalistic study of visitors to a public aquarium.* Unpublished doctoral dissertation, University of California, Berkeley.

Tunnicliffe, S. D. (1998). Boy talk / girl talk: Is it the same at animal exhibits? *International Journal of Science Education, 20*(7), 795–811.

Uzzell, D. (1993). Contrasting psychological perspectives on exhibition evaluation. In S. Bicknell & G. Farmelo (Eds.), *Museum visitor studies in the 90s* (pp. 125–129). London: Science Museum.