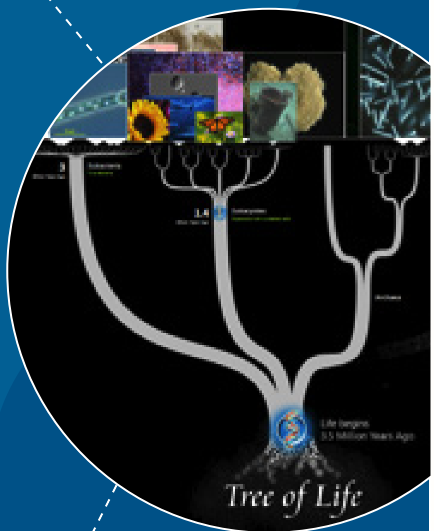



# Life on Earth

## Evaluation Report

James K.L. Hammerman, Ed.D.[1]
Amy N. Spiegel, Ph.D.[2]
Jonathan A. Christiansen[1]

[1]STEM Education Evaluation Center at TERC, Cambridge, MA
[2] Center for Instructional Innovation, University of Nebraska – Lincoln

**STEM Education Evaluation Center at TERC:**

*Improving education through evaluation*

**STEM Education Evaluation Center**
**SEEC at TERC**

TERC

*2067 Massachusetts Avenue*
*Cambridge, Massachusetts 02140*
*617.873.9600 phone*
*617.873.9601 fax*
*http://evaluation.terc.edu*

## Executive Summary

*Life on Earth* is interactive software installed as a museum touchtable exhibit that uses data about over seventy thousand (70,000) species from several databases to help visitors explore and deepen their understanding of biodiversity, evolution and common ancestry, and the history of life on earth (DeepTree/ FloTree). Some installations also include a smaller exhibit that poses puzzle challenges about evolutionary relationships among species (Build-a-Tree (BAT)).[1]

The exhibit was installed at four natural history museums across the U.S. – the Harvard Museum of Natural History (Cambridge, MA), the Field Museum (Chicago, IL), the University of Nebraska State Museum (Lincoln, NE), and the California Academy of Sciences (Cal Academy, San Francisco, CA). Evaluation took place at Cal Academy during two months in the fall, 2012. The *Life on Earth* project also conducted learning research about the impact of the exhibit (Evans et al., 2013; Evans et al., April, 2013).

The project engaged in three inter-related strands of research – Learning Research with youth using experimental methods; human computer interaction (HCI) research on group touch interactions and large data visualization; and this summative evaluation. Together, they build a range of important knowledge about the intervention and its impact. The Learning Research and HCI research are reported elsewhere (Block et al., 2012; Davis et al., 2013; Evans et al., April, 2013). This evaluation was intended to describe how visitors engage with the touch table exhibits when installed in a museum context, including the role of group interaction, and to explore whether engagement with the touch table exhibit helps visitors understand key concepts of evolution. We conducted a video- and audiotaped study of *Life on Earth* exhibit users in a museum context, as well as a naturalistic observation study of exhibit users. A total of 675 visitors were observed using both DeepTree and BAT over the course of over 40 hours during 11 days. A variety of data were collected for the evaluation from both video and naturalistic observations, including time spent, activities engaged in, characteristics of social interactions around the exhibit table, and responses to a short survey about visitors' experiences and their knowledge of evolution and common ancestry.

Based on our observations, the majority of visitors at Cal Academy were white and well educated with substantial museum and technology experience, but there is also significant diversity across age, race/ ethnicity, languages spoken, place of residence, and prior experiences.

The exhibit was designed for groups of visitors to work together with a common focus around the table at the same time, rather than working in parallel. We tried to observe the natural flow of visitors to and away from the exhibit – observing the forming and re-forming of natural groups rather than trying to establish which people came to the museum together – allowing our study to include interactions among strangers. Median group size was 2 or 3 in each exhibit and study condition, with a range from 1 to as many as 7 visitors at a time observed. Visitor groups represented a wide variety of age configurations with more fluid overlap in the naturalistic studies than in the video studies, as expected because of the way we controlled access to the video study.

[1] The *Life on Earth* project is supported by funding from the National Science Foundation (NSF) under award DRL–1010889 (Harvard University; PI: Chia Shen), and this evaluation report is a subcontract to that award. The views expressed herein are those of the authors and may or may not reflect those of the prime awardee or the Foundation.

### Interest and Enjoyment

Visitor self-reports suggest they found both the DeepTree and BAT exhibits interesting and enjoyable and that they would encourage a friend to visit the table (average ratings of 4 or more on a 5 point scale).

These self-reports were backed up by observations, which found that for nearly half of visitors in the DeepTree study, and over 70% in the BAT study, someone verbally expressed enjoyment with use of the software during their time at the exhibit table, and that less than 25% experienced anyone expressing dislike or frustration with the exhibits.

### Engagement

Observations show visitors engaged substantially with the software, spending 1 and $^2/_3$ minutes on average in the naturalistic conditions for both DeepTree and BAT, and about 6 minutes on average in the video condition, controlling for group characteristics.

For the DeepTree video study, where we were able to collect table log data, 38% of visitors engaged with 5 or more (of the 7) major exhibit activities (Top level navigation; Reel item navigation; Inspection of text/ top image zooming; Trait display; Relate; Training tree; and FloTree) and another 22% engaged with 4 of the 7 (total of 59% engaged with 4 or more), suggesting a high proportion of "Diligent Visitors" (Serrell, 1997). Although we were not able to create a reliable multi-dimensional measure of engagement through confirmatory factor analysis, we found that dwell time was the most reliable predictor of a hypothesized underlying latent engagement factor, and dwell times were moderately high.

In addition to these intended types of interactions, evaluators observed visitors engaging in a range of invented behaviors that they used to explore and play with the content of the exhibit. These included "encyclopedic" scanning of species across the canopy of the DeepTree (observing pictures and reading text about one species, then going on to another species); use of manual navigation to back up from the canopy to a common ancestor, then forward down a branch to a related descendant; group attempts to "kill off" one branch in FloTree; and even simultaneous use of an iPad by one teenager to look up information about species while his parents explored the exhibit software, among others.

### Social Interaction

Social interaction around the exhibit table was important. Visitors in all study conditions reported that the presence of others at the table had a somewhat positive impact on their learning and experience at the exhibit, on average.

Between 65% and 90% of visitors in groups of two or more in all software and study conditions experienced verbal negotiation about social interactions. Groups also used a mixture of physical approaches to social interaction, with 85% or more of visitors in the DeepTree video study experiencing turn-taking, two or more people manipulating the exhibit table simultaneously, and visitors pointing to the table without touching it to suggest things to look at or do. (These proportions were somewhat smaller in the DeepTree naturalistic study – 78% experiencing two or more manipulating the table; 50% experiencing pointing; 40% experiencing turn-taking – and similar but slightly larger percentages in the BAT naturalistic study.)

As expected, group configurations were more fluid in the naturalistic study conditions than in the video studies because of increased access control to the exhibit in the video studies. In all study conditions, we found that larger groups tended to spend longer than smaller ones, controlling for other group characteristics, though these effects were only statistically significant for the naturalistic studies – it seems the exhibit design encourages and builds upon positive group interactions. Interestingly, this effect was somewhat tempered in the DeepTree studies if the additional people were "strangers" (i.e., didn't come and go together), suggesting that comfort with social interaction may make an important contribution to this effect.

For the most part, the presence of children and teens did not have a statistically significant impact on exhibit dwell time, though groups containing children or especially younger children tended to spend less time in the DeepTree naturalistic study than did groups containing only adults or those with teens. DeepTree was designed for ages 10 and up, and this dwell time evidence suggests that, as intended, some of the content of DeepTree may be more interesting and engaging for teen and adult learners than for younger children.

The presence of 6-12 year old children (and to a lesser extent, teens) was associated with *increased* time in the BAT studies, though the difference is not statistically significant in the naturalistic study and, though statistically significant in the video study, is based on just 3 groups so may be partially an artifact. Still, it seems the BAT software tended to engage groups with children somewhat more than groups with just adults.

### Biology Content

The content of the exhibits was also important. Visitors reported that they learned moderate amounts from the exhibits (ratings in the mid-3 range on a 5 point scale).

More importantly, analysis of learning outcomes on a subset of the total sample (N=123) suggests that time at the DeepTree exhibit had a statistically significant association with ratings of agreement on the common ancestry questions of the survey and a marginally significant association with agreement ratings on the evolution questions, controlling for visitors' level of education (which is also associated with these scores), and group membership. Engagement in specific activities at the table such as use of the Relate function, use of the FloTree function, or extent of biology talk within the group were *not* associated with differences in these scores. For the BAT exhibit (N=18), educational level is associated with common ancestry scores, but there were no other statistically significant associations.

A similar result was found with the project's own learning research studies – a controlled experiment that used the same questions about common ancestry and evolution. In that study, youth who experienced the DeepTree exhibit were more likely than a control group to agree with the common ancestry questions. In the learning research studies, there was also a positive association between use of the Relate function and increased agreement with common descent. This result was stronger for the younger, 8-11 year-olds, which may be why it was not replicated in the current evaluation, which included adult as well as youth participants in the sample.

Observations show that majorities of visitors in both the DeepTree and BAT naturalistic studies (50% and 65% respectively), and *substantial* majorities in the video studies (over 80%) experienced some talk about biology content while at the exhibit table. This included questions about whether species were related, comments about characteristics of varied species and/ or reading information displayed in the exhibit, and hypotheses about what was going on in the FloTree "Experiment," among others.

In addition, a substantial minority of visitors were able to connect their exhibit experiences to other experiences at the museum and elsewhere in their lives – e.g., learning from classes they had taken, or diseases they had experienced – with about 15% in all study and software conditions expressing one or more such connections during their time at the exhibit, suggesting some integration and meaning-making of the *Life on Earth* experience.

## Conclusions

These findings suggest that the *Life on Earth* exhibit software was successful at engaging a wide range of visitors with a variety of activities to learn about diversity of species, common ancestry, and evolutionary processes. Visitors expressed enjoyment with their experience at the exhibit, and engaged in discussions about biology content and, sometimes, how it related to other experiences in their lives.

Social interactions among visitors around the table were common and larger group sizes at the exhibit were associated with increased dwell time, moreso for groups who came and went together than for those that included “strangers” in the DeepTree studies. Increased dwell time, in turn, was associated with increased common ancestry and evolution scores in the DeepTree exhibit after controlling for visitors’ level of education. This observational study could not make causal inferences directly linking engagement with the DeepTree software with higher scores on important learning outcomes/ goals in the naturalistic setting. However, the evaluation findings align with results from the project’s more controlled experimental learning research which indicates a causal connection for youth between experience with DeepTree and learning results. Together, these findings suggest that exposure to the exhibit software in the museum context may increase scores on important learning outcomes. This is an important set of findings.

# Table of Contents

**Executive Summary** ..... i
Interest and Enjoyment ..... i
Engagement ..... ii
Social Interaction ..... ii
Biology Content ..... iii
Conclusions ..... iii
**Table of Contents** ..... v
**Table of Tables** ..... vii
**Table of Figures** ..... ix

Life on Earth Evaluation Report ..... 1
**Background** ..... 1
Description of Exhibit ..... 1
DeepTree/ FloTree ..... 1
Build-a-Tree (BAT) ..... 3
Learning Goals ..... 3
Case Example ..... 3
**Evaluation Design** ..... 4
Evaluation Research Questions ..... 5
Methods ..... 5
Context and Set-up ..... 6
People – Who's at the table? ..... 9
Events – How do people interact at the table? ..... 10
Inter-Rater Reliability ..... 11
Table Log – What activities do people do? ..... 12
Follow-up Survey ..... 12
Data Collected ..... 12
Data Processing and Coding ..... 13
Analytic Methods ..... 14
**Findings** ..... 15
Description of Visitors by Study ..... 15
DeepTree ..... 15
Build-A-Tree ..... 18
Exhibit Experience ..... 19
Dwell Time ..... 20
DeepTree Exhibit ..... 20
By age, gender, and visit day ..... 21
By group size ..... 23
BAT Exhibit ..... 26
Group Configurations and Interaction Patterns ..... 27
Groupings ..... 27
Group Composition ..... 30
Overlapping of Groups ..... 31
Characterization of groups ..... 32
Parents with children ..... 32
Children alone (not common in the Video Study) ..... 33
Teens ..... 33
Singles ..... 34
Young adults ..... 34
Older adults ..... 34
Other Notes ..... 35

Quantifying Social Interaction Patterns ....................36
DeepTree/ FloTree ....................37
BAT studies ....................40
Table Activity Data ....................44
Qualitative Examination of Table Log Files ....................48
Exhibit Engagement Factor Analysis ....................50
Survey Responses ....................51
Open-Ended Survey Results ....................51
Learning Outcomes ....................52
Relationships Among Variables ....................54
Nested Models for Time Spent ....................54
DeepTree Video Study ....................55
DeepTree Naturalistic Study ....................56
BAT Video Study ....................58
BAT Naturalistic Study ....................58
Nested Models for Learning Outcomes – DeepTree/ FloTree ....................60
Common Ancestry Results ....................61
Evolution Results ....................62
Nested Models for Learning Outcomes – Build-a-Tree (BAT) ....................63
Common Ancestry Results ....................63
Evolution Results ....................63
**Discussion/ Conclusions ....................64**
**Bibliography/ References ....................67**

**Life on Earth Evaluation — Appendices ....................A-i**
**Appendices Table of Contents ....................A-i**
**Table of Appendix Tables and Figures ....................A-ii**
**Study Signage ....................A-1**
**Consent Text ....................A-2**
**Demographic Information ....................A-3**
DeepTree Video Participants ....................A-3
DeepTree Naturalistic Participants ....................A-5
BAT Video Participants ....................A-6
BAT Naturalistic Participants ....................A-8
**Observation Protocol Software ....................A-10**
**Observation Protocol Codebook ....................A-11**
**Observation Protocol Inter-Rater Reliability ....................A-15**
**Dwell Time Details ....................A-17**
By Age Group ....................A-17
By Group Size ....................A-18
**Event Proportions by Intervals ....................A-20**
DeepTree Event Proportions ....................A-20
BAT Event Proportions ....................A-22
**Table Log Summary Measures Definitions ....................A-24**
**Summary Statistics for Table Log Measures ....................A-30**
**Confirmatory Factor Analysis of Exhibit Engagement ....................A-34**
**Semantic Log Visualization Guide ....................A-37**
**Survey Text ....................A-38**
**Survey Results ....................A-41**
DeepTree Survey Results ....................A-41
BAT Survey Results ....................A-43

## Table of Tables


Table 1: Summary of exhibit studies and data collected ....13
Table 2: Self-reported ratings of exhibit experience ....19
Table 3: Time Spent at Table by Age Category ....21
Table 4: Modal Number of People at Exhibit vs. Dwell Time ....25
Table 5: Engagement Behaviors for DeepTree Video Study: Proportion over Total Available Intervals and Proportion of visitors who experienced this behavior ....37
Table 6: Proportion of Visitors who Experienced Engagement Behaviors by Group Size for DeepTree Video Study ....38
Table 7: Engagement Behaviors for DeepTree Naturalistic Study: Proportion over Total Available Intervals; and Proportion of visitors who experienced this behavior ....39
Table 8: Proportion of Visitors who Experienced Engagement Behaviors by Group Size for DeepTree Naturalistic Study ....40
Table 9: Proportion of Engagement Behaviors over Total Available Intervals BAT Video Study ....41
Table 10: Proportion of Visitors who Experienced Engagement Behaviors by Group Size for BAT Video Study ....42
Table 11: Engagement Behaviors for BAT Naturalistic Study: Proportion over Total Available Intervals; and Proportion of visitors who experienced this behavior ....43
Table 12: Proportion of Visitors who Experienced Engagement Behaviors by Group Size for BAT Naturalistic Study ....44
Table 13: Proportion of DeepTree Video Study Visitors Engaging in Different Numbers of Activities....45
Table 14: Selected Table Measure Descriptives ....46
Table 15: Selected Relate Function Descriptives ....47
Table 16: Selected FloTree Descriptives ....48
Table 17: List of Keywords/ Themes for LOE Survey responses ....52
Table 18: Levels of Agreement with Survey Questions about Common Ancestry of Various Species....53
Table 19: Levels of Agreement with Survey Questions about Evolution ....53
Table 20: Final Nested Models Predicting Log-Time at DeepTree Exhibit ....57
Table 21: Final Nested Models Predicting Log-Time at BAT Exhibit ....59
Table 22: Hierarchical Models Predicting Learning Outcomes at DeepTree Exhibit, Video Study ....62


## Appendix Tables


Table 23: DeepTree Video Observation Sample by Age Category and Gender .... A-3
Table 24: DeepTree Video Study Survey Sample by Self-reported Age Category and Gender .... A-3
Table 25: Self-identified Race/Ethnicity of DeepTree Video Observation Survey Sample .... A-4
Table 26: DeepTree Video Participant-reported Education by Age Category .... A-4
Table 27: DeepTree Naturalistic Observation Sample by Age Category and Gender .... A-5
Table 28: DeepTree Naturalistic Study Survey Sample by Self-reported Age Category and Gender .... A-5
Table 29: Self-identified Race/Ethnicity of DeepTree Naturalistic Survey Sample .... A-5
Table 30: DeepTree Naturalistic Participant-reported Education by Age Category .... A-6
Table 31: BAT Video Study Sample by Age Category and Gender .... A-6
Table 32: BAT Video Study Survey Sample by Self-reported Age Category and Gender .... A-6
Table 33: BAT Video study Self-identified Race and Ethnicity of Survey Sample .... A-7
Table 34: BAT Video Study Participant-reported education by Age Category .... A-7
Table 35: BAT Naturalistic Study Sample by Age Category and Gender .... A-8
Table 36: BAT Naturalistic Study Survey Sample by Self-reported Age Category and Gender .... A-8
Table 37: BAT Naturalistic Study Self-identified Race and Ethnicity of Survey Sample .... A-8
Table 38: BAT Naturalistic Study Participant-reported education by Age Category .... A-9
Table 39: Time Spent at Table by Age Category for DeepTree Video study .... A-17
Table 40: Time Spent at Table by Age Category for DeepTree Naturalistic study .... A-17
Table 41: Time Spent at Table by Age Category for BAT Video study .... A-17

Table 42: Time Spent at Table by Age Category for BAT Naturalistic study ...................................... A-18
Table 43: Modal Number of People at Exhibit vs. Dwell Time for DeepTree Video Study ................. A-18
Table 44: Number of People at Exhibit vs. Dwell Time for DeepTree Naturalistic study..................... A-18
Table 45: Number of People at Exhibit vs. Dwell Time for BAT Video Study.................................... A-19
Table 46: Number of People at Exhibit vs. Dwell Time for BAT Naturalistic Study............................ A-19
Table 47: Proportion of Engagement Behaviors (SD) over Total Available Time Intervals by Group Size for DeepTree Video Study ........................................................................................... A-20
Table 48: Proportion of Engagement Behaviors over Total Available Intervals by Group Size for DeepTree Naturalistic Study ........................................................................................... A-21
Table 49: Proportion of Engagement Behaviors over Total Available Intervals by Group Size for BAT Video Study........................................................................................................... A-22
Table 50: Proportion of Engagement Behaviors over Total Available Intervals by Group Size BAT Naturalistic Study.................................................................................................. A-23
Table 51: General Table Activity Descriptives ..................................................................................... A-30
Table 52: Navigation Activity Descriptives .......................................................................................... A-30
Table 53: Inspection Activity Descriptives ........................................................................................... A-31
Table 54: Trait Activity Descriptives .................................................................................................... A-31
Table 55: Relate Function Descriptives................................................................................................. A-32
Table 56: Training Tree Activity Descriptives ...................................................................................... A-32
Table 57: FloTree Descriptives.............................................................................................................. A-33
Table 58: List of Indicators for Confirmatory Factor Analysis ............................................................ A-34
Table 59: LOE Indicator Reliabilities (Item Reliability Estimates) ...................................................... A-35
Table 60: Indicator-factor Correlation (Validity Estimates).................................................................. A-35
Table 61: Self-reported ratings of exhibit experience, DeepTree Video Study..................................... A-41
Table 62: Levels of Agreement with Survey Questions about Common Ancestry of Various Species. DeepTree Video Study ............................................................................. A-41
Table 63: Levels of Agreement with Survey Questions about Evolution. DeepTree Video Study ....... A-41
Table 64: Self-reported ratings of exhibit experience, DeepTree Naturalistic Study............................ A-42
Table 65: Levels of Agreement with Survey Questions about Common Ancestry of Various Species. DeepTree Naturalistic Study ...................................................................... A-42
Table 66: Levels of Agreement with Survey Questions about Evolution. DeepTree Naturalistic Study ........................................................................................... A-42
Table 67: Self-reported ratings of exhibit experience, BAT Video Study ............................................ A-43
Table 68: Levels of Agreement with Survey Questions about Common Ancestry of Various Species. BAT Video Study........................................................................................................... A-43
Table 69: Levels of Agreement with Survey Questions about Evolution. BAT Video Study ............... A-43
Table 70: Self-reported ratings of exhibit experience, BAT Naturalistic Study ................................... A-44
Table 71: Levels of Agreement with Survey Questions about Common Ancestry of Various Species. BAT Naturalistic Study.................................................................................................. A-44
Table 72: Levels of Agreement with Survey Questions about Evolution. BAT Naturalistic Study ...... A-44

## Table of Figures

Figure 1: Screen Shots of DeepTree/ FloTree Software ..... 2
Figure 2: Screen Shot of Build-a-Tree Software ..... 3
Figure 3: California Academy of Sciences East Wing Exhibit Layout ..... 7
Figure 4: Evaluation Video Installation ..... 8
Figure 5: Schematic of Evaluation Video Installation ..... 9
Figure 6: DeepTree Age Distributions ..... 16
Figure 7: Build-A-Tree Study Age Distributions ..... 18
Figure 8: Graph of Time Spent at Table by Age Category for DeepTree Video Study. ..... 22
Figure 9: Graph of Time Spent at Table by Age Category for DeepTree Naturalistic Study. ..... 23
Figure 10: Modal Number of People at Exhibit vs. Dwell Time for DeepTree Video Study ..... 24
Figure 11: Modal Number of People at Exhibit vs. Dwell Time for DeepTree Naturalistic Study ..... 25
Figure 12: Modal Number of People at Exhibit vs. Dwell Time for BAT Video Study ..... 26
Figure 13: Modal Number of People at Exhibit vs. Dwell Time for BAT Naturalistic Study ..... 26
Figure 14: DeepTree Video Study Group Connections ..... 28
Figure 15: DeepTree Naturalistic Study Group Connections ..... 29
Figure 16: BAT Video Study Group Connections ..... 30
Figure 17: BAT Naturalistic Study Group Connections ..... 30
Figure 18: Log file visualization: User 297 ..... 49
Figure 19: Log file visualization: User 238 ..... 50

## Appendix Figures

Figure 20: Screen Shot of Observation Protocol Database ..... A-10
Figure 21: CFA – Table Engagement: Factor Loadings ..... A-35

# Life on Earth Evaluation Report

"*Life on Earth*" is an NSF-funded project[2] that aims to (1) advance public understanding of biodiversity and the history of life on Earth, and (2) advance our knowledge of how people interact with and learn from large interactive science data visualizations on multi-touch tabletop displays in public settings.[3] The project allows museum visitors to explore evolutionary concepts through unique and interactive visualizations of the phylogenetic tree of life and the process of speciation. The project accomplishes this through the use of an interactive touch table computer in which museum visitors can explore the tree of life by navigating through it, exploring how two species relate or finding species within the tree. Visitors can also use an experiment function to explore the process of speciation by creating barriers with their hands to create speciation events. The tree of life integrates several databases into a single interactive exhibit. These include: Tree of Life web project, Encyclopedia of Life, National Center for Biotechnology Information and Time Tree.

The *Life on Earth* project is led by Principal Investigator (PI) Chia Shen of the Harvard University School of Engineering and Applied Sciences (SEAS), Cambridge, MA, and Co-PIs Judy Diamond, University of Nebraska, Lincoln, NE, E. Margaret Evans, University of Michigan, Ann Arbor, MI, and Michael Horn, Northwestern University, Evanston, IL. The summative evaluation of the project[4] was carried out by James K. L. Hammerman and Jonathan A. Christiansen from the STEM Education Evaluation Center (SEEC) at TERC, a non-profit education research and development organization in Cambridge, MA; and Amy N. Spiegel from the Center for Instructional Innovation, University of Nebraska – Lincoln. Heather A. Lavigne, of the University of Massachusetts, Amherst, provided quantitative analysis support. This report provides summative findings of the project based on two sets of studies that were carried out from October 2012 through December 2012 at the California Academy of Sciences, San Francisco, CA. Data were collected by Anita Smith, principal of Mountain Light Consulting, Sebastopol, CA, assisted by Julie Shattuck of Shattuck Applied Research and Evaluation, Santa Cruz, CA, and Lauren Hodge, independent evaluator.

## Background

The touch table exhibit was installed in four museums across the United States: Harvard Museum of Natural History in Cambridge, MA; University of Nebraska State Museum in Lincoln, NE; California Academy of Sciences in San Francisco, CA, and the Field Museum in Chicago, IL. The exhibit consisted of two independent pieces of software, *DeepTree/ FloTree (DeepTree)*, and *Build-a-Tree (BAT)*, both of which are described in more detail below.

### *Description of Exhibit*

#### DeepTree/ FloTree

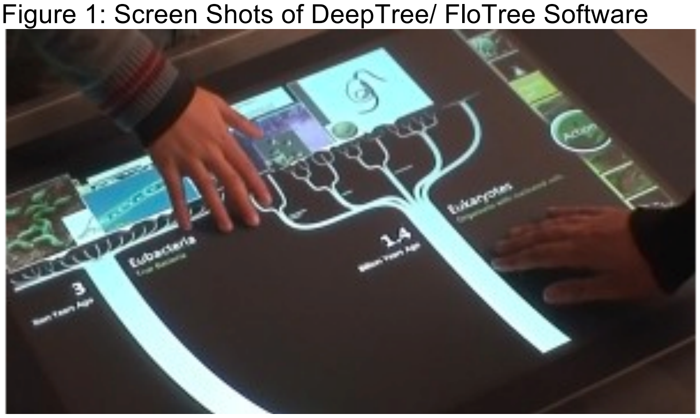
The focus of this study, the *Tree of Life* exhibit, incorporates multiple components, each of which provides specific interactive learning modes on the touch surface. The overarching

[2] National Science Foundation support for this work is gratefully acknowledged under a subcontract to award DRL–1010889 (Harvard University; PI: Chia Shen). The views expressed herein are those of the authors and may or may not reflect those of the prime awardee or the Foundation.

[3] https://lifeonearth.seas.harvard.edu/

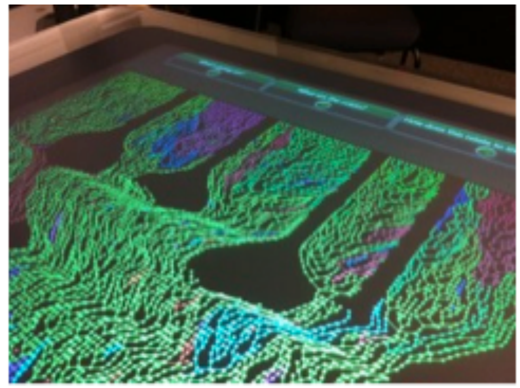
structure is the DeepTree, a tree visualization of the tree of life that allows both free exploration and several specific entry points for deeper investigation and learning activities (see Block et al., 2012; Davis et al., 2013 for more detail on the DeepTree design). The DeepTree exhibit shows the ancestral relationships of 70,000 species, both living and extinct, dating back to the origins of life 3.5 billion years ago. In the DeepTree, users can navigate through the tree manually as well as by holding down an image of an organism displayed on the canopy, which causes the display to "fly" through the tree to the selected species, where more information about the selected species is provided.

Figure 1: Screen Shots of DeepTree/ FloTree Software




DeepTree

For a video demonstration, go to: https://lifeonearth.seas.harvard.edu/learning-activities/deeptree/



FloTree

For a video demonstration, go to: https://lifeonearth.seas.harvard.edu/learning-activities/flotree/

A second component is a scrolling reel along the right side of the screen with images representing 200 pre-selected species across a wide range of organisms. Users can scroll through and select an image from this reel. When a visitor pulls an image onto the main display and holds it, a transparent chord pointing to the species' location is shown while the display again "flies" to that location. As the display automatically "flies" through the tree, all the intervening nodes and branches become momentarily visible.

A third component is an action button, also on the right hand side and partially overlapping the image reel in the center. This button reveals three options when tapped: *Relate*, *Experiment* and *Return*. *Relate* allows the user to select any two species from the reel and the display zooms in to their most recent common ancestor, with their shared lineage highlighted in the tree. The user is then prompted to tap on an icon that reveals the "*training tree,*" a simplified tree showing the species' shared lineage and identifying major speciation nodes. These nodes can be tapped to open a "*trait display*" that provides more detailed information. This trait display can also be activated from a limited number of active nodes on the DeepTree.

*Experiment* opens FloTree, a multi-user simulation activity designed to illustrate population processes related to evolutionary change, in particular, the process of speciation through physical separation of genetically varying subpopulations (see Chua et al., In Press 2013 for more detail on FloTree design). As users watch successive "generations" of animated dots that "grow" upward, they can put their hands or arms on the surface to divide the single population into subpopulations separated by "physical" barriers. As these subpopulations evolve independently from one another, they eventually lead to the development of separate species in the simulation, and when the simulation finishes, the resulting phylogenetic tree can be displayed. During the evaluation data collection, additional text explanations about the simulation were available as options for display, but these were later incorporated as permanent visual elements included as part of the simulation activity.

### Build-a-Tree (BAT)

A second exhibit installed on the table was *Build-a-Tree (BAT)*, an interactive, multi-level puzzle game designed to help visitors learn about evolution and common ancestry. Visitors are guided in the construction of trees that show the evolutionary relationships among different organisms. As visitors correctly construct the subtrees, labels of shared traits are displayed. The several levels of this game become more complex as visitors successfully construct each tree composed of an increasing number of pre-selected organisms. Prior evaluation on this exhibit indicated that visitors were engaged with the game, enjoyed collaborating, and were successful in moving through the levels of the game (Horn et al., 2012). We include a small evaluation of BAT in this summative report.

Figure 2: Screen Shot of Build-a-Tree Software

Note: For a video demonstration, go to https://lifeonearth.seas.harvard.edu/learning-activities/bat/

## ***Learning Goals***

Five learning goals guided the overall design of the exhibits:

1. all life on Earth is related;
2. biodiversity is vast;
3. relatedness is derived from common descent;
4. species inherit shared traits from common ancestors; and
5. evolution is ongoing and happens over very long periods of time (Davis, et al 2013)

## ***Case Example***

We begin with an example of how visitors engage with the DeepTree/ FloTree exhibit. This is not "typical" but demonstrates how a cooperative, interested, and somewhat knowledgeable adult pair experience all three main activities at the table:

> *One twenty-something couple (man and woman) spent a long time, about 25 minutes, exploring the table. They started by pulling up the video on DNA, but it was so loud at the museum, they weren't even sure there was audio, and asked the evaluation observers about it. (Near the end of their time at the exhibit, they accessed that screen and audio again, and appeared able to hear it then.)*
>
> *As they began to manually explore the DeepTree, they read some of the text aloud (names of species, number of years ago), and then while the man drove, the woman remarked that she did "not like watching with you controlling it." She then did the driving*

*for awhile. She later said it made her dizzy and couldn't watch the tree zooming by. They took turns driving and exploring the DeepTree manually in a very cooperative manner. As they were browsing through the DeepTree, he said, "I kind of feel lost," and she agreed.*

*Once they got to the whole tree with the root visible, he started to navigate up the tree in search of humans. "We were only 2 million years ago, right?"*

*She replied, "Hmmm-hmm. So you're all the way back to 850 million years ago."*

*He said, "Yeah, but we're getting closer as we get up the chain."*

*They took a detour to look at some of the jellyfish, and expanded the photos to see them close-up. They took turns to move through the tree, and were persistent and methodical in exploring and trying to find specific species.*

*Finally, after closing the right hand trait window that had been open the whole time during their manual exploration, they saw the reel of organisms. They continued to manually explore, reading species' names as he drove.*

*She then said, "It would be kind of cool if you could click on a word and have them define it, for those of us who have no biology background…like 'what are the characteristics of a mammal?" After a few attempts at tapping on the words around one of the tree nodes, the right hand window opened again with the related information, and she was pleased, "That's cool."*

*Later, as they were moving through the DeepTree, she said, "Humans think they are so special, but they are like such a tiny piece."*

*As he manually moved back in time through the tree, watching the branches, she asked, "Are you excited by this?" and he replied, "I really am."*

*They eventually used the buttons on the right to more quickly navigate the tree, and also eventually pulled species from the reel.*

*They did the FloTree experiment twice (it crashed once), and when they used Relate, they remarked, "That's so cool."*

*"So 1.4 billion years ago they were connected."*

*"That's really cool."*

This couple discusses several important aspects of biological content, seeking specific information from the exhibit and usually being able to find it. They work together to navigate the DeepTree in a variety of ways, engaging with the exhibit and its content and taking pleasure in their experience. Although the length and depth of their experience is not typical, these important aspects of their experience are not unusual.

## Evaluation Design

The design of the evaluation for this exhibit drew upon both traditional museum visitor studies focused on measures such as dwell time at an exhibit to assess engagement and associated learning (Borun, Chambers, & Cleghorn, 1996; Serrell, 1997), and more recent studies on visitor learning at evolution exhibits (Evans et al., 2010; Spiegel et al., 2012). In addition, recent work on interactive tabletop applications contributed to the kinds of questions guiding the evaluation. Interactive tabletop exhibits are becoming increasingly common in informal learning settings as an innovative way to provide hands-on experiences (e.g., Lindemann-Matthies & Kramer, 2006; vom Lehn & Heath, 2005), and with multi-touch surfaces, as a way for multiple visitors to interact and work together. Studies have shown that multi-touch surfaces support learning and collaboration among students (Harris et al., 2009; Higgins, Mercier, Burd, & Joyce-Gibbons, 2012; Rick, Rogers, Haig, & Yuill, 2009), and can encourage playfulness (Jacucci et al., 2010). Observations and self-report measures indicate users enjoy the novelty and interactivity of the touch table surface (Hornecker, 2008; Jacucci et al., 2010), but assessing other impacts can be challenging, particularly in free choice environments. One evaluation study looking at another interactive *Tree of Life* tabletop exhibit found that some visitors, although they enjoyed their experiences at the table, were unable to decipher the intended content of the exhibit (Hornecker, 2008).

The findings in these studies highlight the important fact that user interaction with these devices is dependent on the specific application installed on the table. Particularly with respect to social interaction around the table, the design of the exhibit and identified learning goals provide the unique setting for the visitor experience. Two interactive exhibit models guiding the *Tree of Life* exhibit were Planned Discovery (PD) and Active Prolonged Engagement (APE) (see Humphrey, Gutwill, & The Exploratorium APE Team, 2005). PD exhibits, now common in hands-on science museums, present information in a mostly prescriptive style, with specific instruction designed to make the focal phenomena accessible. APE design represents a shift toward more exploratory, visitor-directed, and open-ended interactive exhibits. In addition, the *Life on Earth* exhibit was designed to encourage social interaction around the content of the exhibit. Consequently, how visitors actually interacted with the table and with other visitors while at the table was a particular focus for this evaluation.

## Evaluation Research Questions

The *Life on Earth* Exhibit summative evaluation addressed the following key questions to more fully understand the impact of the project:

1. *How do visitors engage with the touch table exhibits, when installed in a museum context?*
   What is the nature of the engagement with the exhibits? How long is the engagement, what kinds of activities does it consist of, and how does this vary across different kinds of visitors?
2. *To what extent does the exhibit support group interaction and social play?*
   How do people interact with the exhibit, both individually and as a group?
3. *To what extent does engagement with the touch table exhibit help visitors understand key concepts of evolution?*
   In other words, what is the nature of the learning taking place as a result of interacting with the exhibit? Although this question was primarily addressed by a separate learning research team, we incorporated a subset of their instruments/ items to correlate results with data about engagement.

## Methods

To answer these evaluation research questions, the summative evaluation team carried out two studies of DeepTree/ FloTree and a smaller pair of studies of Build-a-Tree (BAT), both at the California Academy of Sciences from October 2012 to December 2012. The first study (Video) was a video and audiotaped study of *Life on Earth* exhibit users in a museum context; the second study (Naturalistic) was a naturalistic observation study of exhibit users. The video study provided more detailed data about visitor interactions with the table and its impact on their knowledge than the naturalistic study; but the process of signing consents and entering a restricted area may also have modified behavior from what it would have been naturally. Studying behavior in both contexts allows us to better understand likely outcomes for the unstudied museum visitor. Piloting for the studies and instruments was done throughout the summer and fall of 2012 at the Harvard Museum of Natural History.

## Context and Set-up

Located in Golden Gate Park in San Francisco, CA, the California Academy of Sciences (Cal Academy) is "a multifaceted scientific institution committed to leading-edge research, to educational outreach, and to finding new and innovative ways to engage and inspire the public."[5] Housed in a 412,000 square foot structure, the space includes an aquarium, a planetarium, a four-story glass terrarium, a natural history museum, and a variety of other exhibits, as well as research and exhibit design facilities, teaching spaces, a restaurant, and gift shop, among others. The *Life on Earth* exhibit was installed on the main floor of Cal Academy, in the East Wing of the building, near a variety of exhibits on biology and evolution (e.g., Darwin's finches, Baobab exhibit, Bug rug, Insect collecting game, Island colonization, Tortoise wall) and on current scientific work (e.g., Living lab and Science in action) (Figure 3). The space has ceilings several stories tall, concrete floors, floor to ceiling windows along the entire east wall, and can be quite loud when occupied by excited visitors. In order to reduce extraneous light that could affect functioning of the touch table, the museum staff crafted a vertical plywood board installed between the table and the windows about 3 feet from the table; and an 8 foot high x 12 foot wide x 11 foot deep overhanging ceiling made of wire frame panels covered in black mesh. A 10 inch x approximately 8 foot light panel hung at the front of the ceiling panels with the words "TREE OF LIFE" (Figure 4). This set up was repeated twice – once on the left for the touch table running Build-a-Tree software; once on the right for the table running DeepTree/ FloTree software. There was a triangular pylon between the tables printed with exhibit signage on the outside, that housed the evaluation recording equipment and its power supply inside.

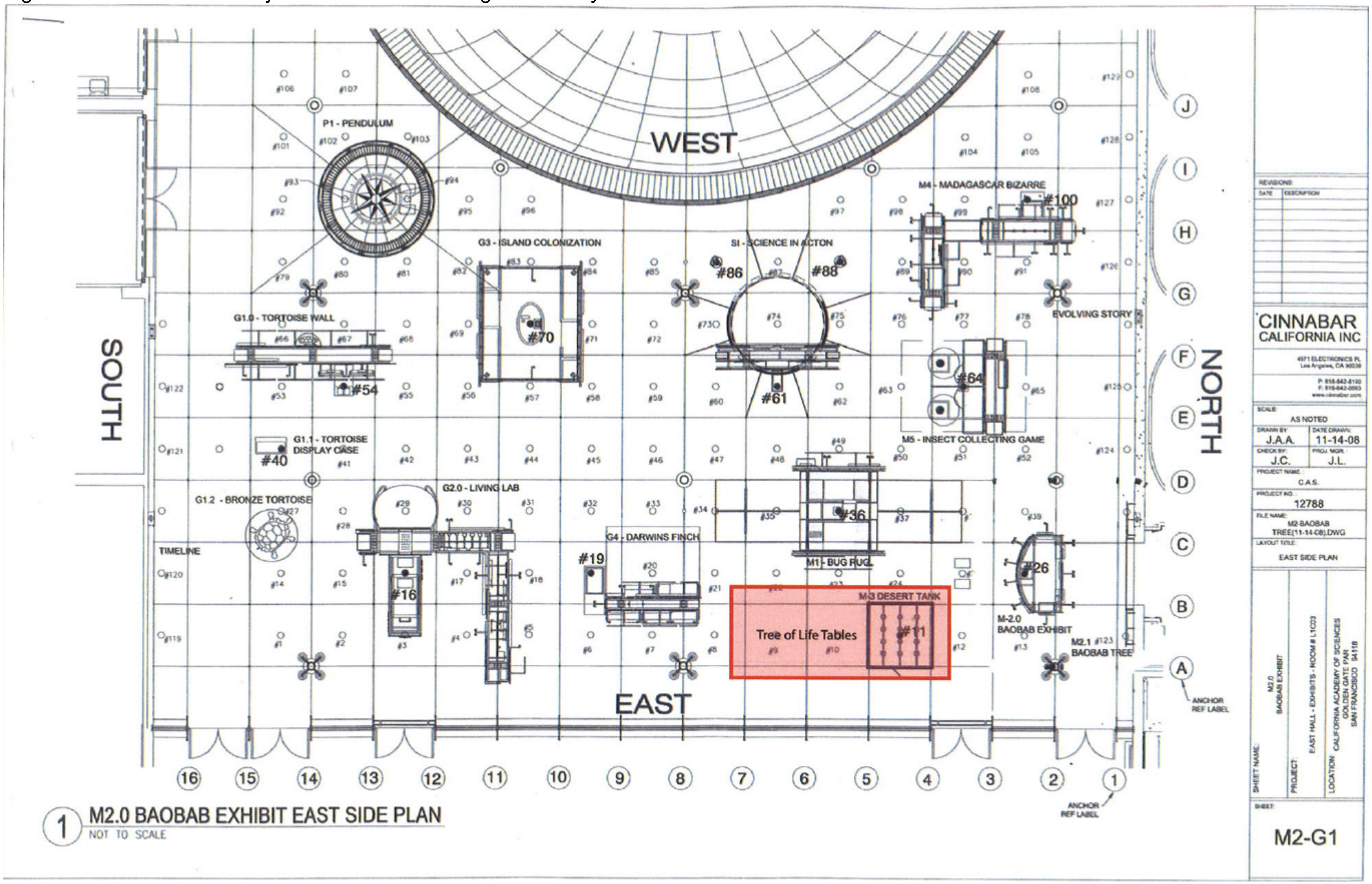
While the backboard and ceiling mesh installations improved the functioning of the exhibit, they increased some of the existing challenges for the evaluation research, primarily with respect to capturing high quality sound for the video study. Before arriving we were concerned about ambient noise at the museum and, once on site, found that the volume of sound from nearby exhibits – particularly the repeating audio track attractor from "Science in Action" exhibit about a woman scientist studying parasites; and from an interactive exhibit about insects on the forest floor ("Bug Rug") that seemed to encourage children to "stomp" on them – were particularly problematic. At the same time, the solid wood backboard reflected some of this sound back into the *Life on Earth* exhibit area; and the mesh ceiling panels prevented us from installing overhead mikes any closer than 5.5 feet above the surface of the exhibit table. The evaluation equipment installation is detailed below and diagrammed in Figure 5.

The Video and Naturalistic studies were carried out by three observers, trained by SEEC evaluator Hammerman on site at Cal Academy during several days in October, 2012, with follow-up work to establish inter-rater reliability for event coding conducted during several subsequent weeks.

For the Video study, observers cordoned off an area around the touch table (Figure 4), controlled entry and egress with explicit informed consent for participation in the research. Museum visitors were invited by one member of the observation team to use the table, and completed informed consent documentation describing the study and its methods, including use of video and audio recording, before entering the exhibit space. Written parent/ guardian consent was required for children and teens (see Appendices for text of signage announcing the study, Study Signage, p. A-1, and text of informed consent documents, Consent Text, p. A-2). Each individual was then given a unique ID sticker to place on themselves for identification purposes. While in the vicinity of the touch table exhibit, museum visitors were video and audio recorded.

[5] http://www.calacademy.org/academy/about

Figure 3: California Academy of Sciences East Wing Exhibit Layout



A second observer monitored the audio and video equipment through a set of headphones, and coded visitor behavior and talk using an observation protocol described below. A third observer administered surveys to visitors as they voluntarily left the study area.

For the Naturalistic study, posted signs near the exhibit described that research observations were taking place, and more detailed information sheets were available for those who were interested (see Appendices, Study Signage, p. A-1 and Consent Text, p. A-2), but the area was freely accessible to all visitors. The study was staffed by two or three observers: one or two coding for people's arrival and departure, and human interactions around the table (separately if two observers; one person covering both if only one observer); the second or third observer administering surveys to visitors exiting the area, when they agreed to participate, and answering questions about the study as necessary.

Figure 4: Evaluation Video Installation

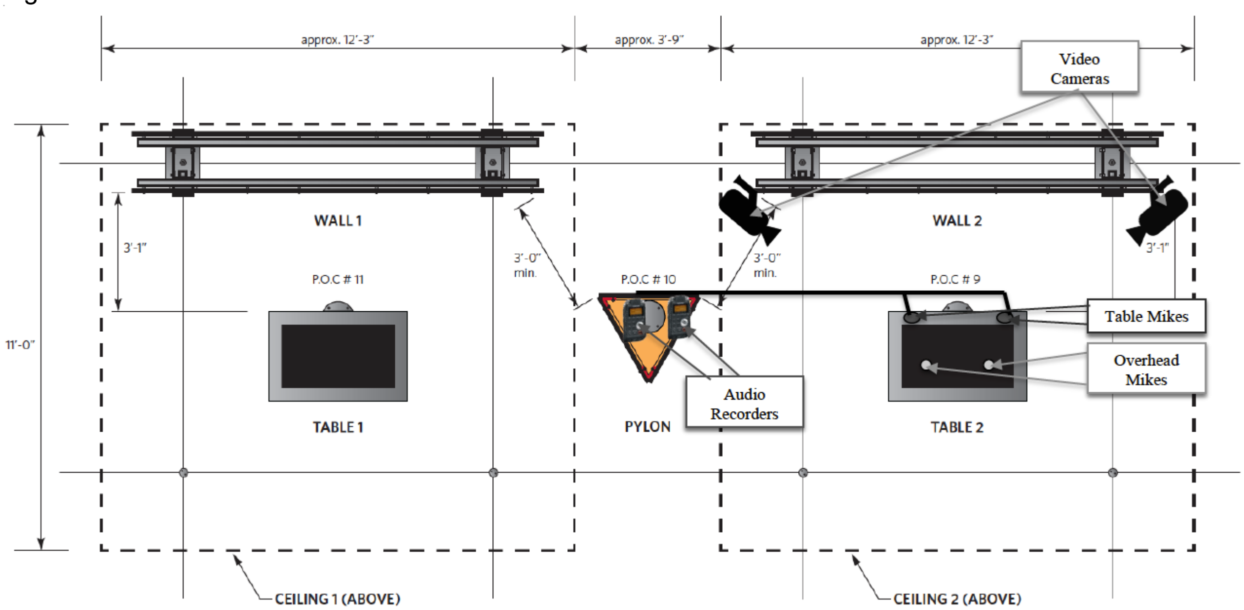

Recording equipment for the Video study consisted of two overhead cameras with mikes positioned at the upper back corners of the exhibit and aimed to capture behavior at the table but not beyond the cordoned area. Capturing sound with sufficient fidelity to code visitor talk was important and difficult in the relatively noisy Cal Academy space. Positioning mikes as close to people's voices as possible was key – we used two flat table mikes taped to the back corners of the table, and two overhead mikes positioned directly over the table, above the mesh canopy that shaded the table from stray light. Each pair of mikes led to a separate digital field recorder, housed in the triangular pylon to the left of the exhibit area that also served as signage for the exhibit (see A second observer monitored the audio and video equipment through a set of headphones, and coded visitor behavior and talk using an observation protocol described below. A third observer administered surveys to visitors as they voluntarily left the study area.

For the Naturalistic study, posted signs near the exhibit described that research observations were taking place, and more detailed information sheets were available for those who were interested (see Appendices, Study Signage, p. A-1 and Consent Text, p. A-2), but the area was freely accessible to all visitors. The study was staffed by two or three observers: one or two coding for people's arrival and departure, and human interactions around the table (separately if two observers; one person covering both if only one observer); the second or third observer administering surveys to visitors exiting the area, when they agreed to participate, and answering questions about the study as necessary.

Figure 4: Evaluation Video Installation

and Figure 5 for photo and diagram of installed equipment). Microphones on the cameras were used primarily to provide an audio track used for synchronizing the video images with the higher quality sound recorded from the table and overhead mikes. Cameras and mikes were installed and removed each day the study occurred. Data were downloaded from video and audio cards and backed up to two portable hard drives, and field notes were sent to the SEEC evaluation team daily.

Figure 5: Schematic of Evaluation Video Installation




## People – Who's at the table?

The observers coded visitor talk and behaviors directly into a Filemaker (Filemaker Inc., 2011, 2013) database designed by SEEC evaluators running on an iPad (see Appendix Observation Protocol Software, p. A-10 for observation protocol). Although Filemaker allows for direct interactive recording of data on an external server, the constraints of wifi and internet connection speeds at Cal Academy made this impractical for interactive use – a constraint which is likely to be common in museum contexts. Instead, data were recorded locally on iPads and files were sent to TERC at the end of each day.

To track dwell time and number of people at the table, observers pressed a button on the database tool to automatically record when people arrived and left the table, as well as coding basic demographic characteristics (gender and age category, to the extent we could discern it). They also coded for 19 different behaviors and interactions between people and the table, and among the people at the table (described under Events – How do people interact at the table?, below).

Although we tracked each individual's dwell time, the unit of analysis for behavioral "event" observations was the group who happened to be around the table at any one time. We coded for group behaviors because we assumed that engaging with the exhibit could occur not only by manipulating the software features oneself, but also by watching others manipulate the table, and/or interacting with others through words and gestures (see also Serrell and Associates, 2009). In practice, this meant that we didn't note who at the table said or did something, just that it occurred while the group was around the table.

We also coded for natural groups rather than pre-identifying groups who arrived and left together. Even in the Video study, we didn't control who had access to the table (beyond the consenting process), so that group interactions could better reflect the free flow arrival and departure of visitors that typifies naturalistic interactions with the exhibit. (However, we did notice from the video that occasionally our evaluation observers would close off the area when a group was at the table.) This means that the "group" around the table at any one time may or may not know one another prior to interacting at the exhibit. In the analysis section below, we describe how we characterize "group size" given this free flow of visitors and try to account for common experience in groups through use of hierarchical modeling.

### Events – How do people interact at the table?

Because we were coding for so many different types of behaviors and interactions, we also didn't code *every* time that a behavioral event occurred. Instead, interactions were coded in twenty (20) second intervals – noting that an identified behavior occurred during an interval; and noting it again if it occurred in a subsequent interval. We refer to coded behaviors that occur within a 20 second time interval as "events" –something that the whole group experiences collectively. While this method loses exact counts of the number of certain types of interactions, it allows us to look for and code less-frequent interactions when they occur, instead of having them be swamped by the constant recording of frequent ones (such as table touches). Number of intervals in which a behavior occurs serves as a proxy for frequency/ extent to which that behavior occurs – but actually are minimum counts because behaviors could have occurred more than once in the interval. During pilot testing, we experimented with different interval lengths, starting with 15 second intervals, and eventually deciding that a 20 second interval allowed for more accurate and reliable coding.

Working with input from project staff, we developed and refined the coding scheme over several months, weighing importance for understanding interactions around the exhibit and visitor experiences, degree of interpretation needed for reliable coding, and overall coder burden. We further refined our codebook while we were establishing reliability, in order to clarify the meaning of codes and how to implement them in the field. The final nineteen (19) coded interactions were grouped into four rough categories:

1. "People Gestures/ Touch" interactions which described physical interactions among people around the table such as helping a child by lifting them or moving their hand, blocking others' access to areas of the table or otherwise preventing them from

touching the table, pulling someone towards the table, or yielding to someone else at the table (this last type turned out to be difficult to code).

2. "Table Gestures/ Touch" interactions which described how people touched the table – one at a time, two at a time, three or more at a time; by taking turns (a sequence of different people touching one at a time, which could also occur across coding intervals), or by pointing at places on the table or things to do which, though not literally touching the table, was a physical interaction primarily between a person and the exhibit table. It was important for us to code number of people touching the table at once since the software itself couldn't distinguish between a single person touching in several places, or several people touching at once.
3. "Verbal Interactions" which describe things people say to one another around the exhibit. Specifically, we coded for biology questions and statements, comments about how to use the table or what to do with it, comments that constituted social negotiation about what to do with the exhibit or whether or not to leave the exhibit area, times when people seemed to be reading text displayed on the exhibit, and references to other experiences at the table or elsewhere as that suggested people integrating their experience at the exhibit with other things they knew. Finally, we coded for when people were talking with one another but it was unintelligible either because their voices were too soft to overcome the ambient noise level, or because they were speaking in a language observers didn't understand.
4. "Emotional Expressions" we noted when people expressed positive enjoyment of the exhibit, or some kind of dislike, frustration or negative feeling about the exhibit.

Observers could also take brief open notes about what they observed at the table, which were mostly used to cue more extensive reflective note-taking after an observation session was complete. A screen shot of the database tool, and our code book are in the Appendices (Observation Protocol Software, p. A-10, and *Observation Protocol Codebook*, p. A-11.)

Though our observers coded for behavioral Event data in the field for the DeepTree Video study, this was during the period when we were establishing inter-rater reliability, so final Event data for the Video study came from observation and coding from the videotape by Spiegel and Hammerman several months later. However, observations in the field served as a check on what was recorded by the video, and pointed us towards a few cases where one of our two cameras failed, but the other captured interactions we hadn't yet processed.

*Inter-Rater Reliability*

In order to feel confident about our observations as we proceeded to the Naturalistic study, we spent several weeks establishing inter-rater reliability among our field observers and evaluation staff. Inter-rater reliability for Event coded data was established using video gathered at the Harvard Museum of Natural History during the pilot testing phase, and from a test video created at Cal Academy during observer training. Though average percent agreement was 90% among the five evaluators raters (Jim, Amy, Jon, Anita, and Julie), Cohen's Kappa statistic which accounts for chance agreement was only fair to moderate ($\kappa = .40$) for all codes, and slightly higher ($\kappa = .45$) when limited to codes whose prevalence was between 10% and 90% (Hallgren, 2012). Upon reflection, we realized that our time-driven coding system may have been leading to deflated Kappa values, as the same behavior was coded in successive intervals by different raters. We created a modified "Next Kappa" statistic to account for this, which yielded much higher values ($\kappa_{Next} = .68$). For codes which had high or low prevalence, Byrt, Bishop & Carlin's Kappa correction (1993) cited in (Hallgren, 2012) gave values $\geq .94$. Together, these statistics seemed sufficiently high to gave us confidence in moving to the naturalistic study for which there would be no video back-up data. Rechecking our reliability as we later coded the video data, agreement was somewhat smaller but still acceptable (average $\kappa_{Next} = .62$). See Appendix (Observation Protocol Inter-Rater Reliability, p. A-15) for details about our process of establishing inter-rater reliability.

### Table Log – What activities do people do?

In addition to the videotaped record of people's interactions around the *Life on Earth* exhibit, and our observers' coding of human interactions, the touch table software itself was programmed to log 42 actions – the beginning and ending of software processes, times of visitor interactive touches at the table, or (largely) markers of locations in the tree of life that were displayed as visitors navigated through the exhibit. Evaluators worked with the development team during the spring, 2013, to transform these raw action data (often over 100,000 records for a several hour session) into 78 higher order measures of visitor behavior or access to information or data from the tree of life (see Appendix, Table Log Summary Measures Definitions, pp. A-24 to A-30). Measures characterize number of times people engaged with different portions of the software environment or accomplished various actions (e.g., successfully reaching the target species for a reel-item search, or conducting a FloTree "experiment" that results in one or more speciation events), amount of time spent, and extent of the DeepTree viewed while navigating in different ways, among other things.

Our intention was to synchronize the Table Log data with the data coded by observers in order to have a complete picture of what they were doing, and how they were interacting with others around the exhibit. However we ran into a variety of problems synchronizing these data sources, so we were only able to link these several data sources for the DeepTree Video study. More on this process is described below under analysis.

### Follow-up Survey

Visitors were allowed to stay in the exhibit area for as long as they liked. Once they left, they were invited to complete a follow-up survey that posed an open-ended query eliciting visitors' view of the content of the exhibit; questions about their experience at the exhibit focusing on interest, enjoyment, learning, and the impact of working with others; a few questions taken directly from the Learning Research designed to tap important understandings about common ancestry and evolution; and a variety of demographic and background questions. Visitors in the Video study recorded their Sticker ID (often just by attaching their sticker to their survey) so that we could link behaviors around the touch table exhibit with results from the survey. For those in the Naturalistic study, the observer recorded the time the visitor left the exhibit area for rough coordination with other data. A copy of the Survey appears as Appendix, Survey Text, p. A-38.

### Data Collected

In the DeepTree Video study, visitor interaction with the exhibit was video- and audio-recorded in a series of sessions. Six days of video/audio on the DeepTree/FloTree exhibit (October 20, 22, 28, 30; and November 2 & 8, 2012) totaling 16 hours, 27 minutes were recorded, and observations of visitors were coded from the video records. As described earlier, when visitors are aware that they are being recorded and must go through an informed consent process to get access to the touch table exhibit, they may not behave as naturally as they would in a more typical museum exhibit setting. Thus, we also collected more naturalistic data that was not video- or audio-taped. In the naturalistic study, observations of visitor interactions with the exhibit which occurred over three days (November 27; and December 6 and 9, 2012) totaling 14 hours, 31 minutes, were coded in real time (DeepTree Naturalistic study). The videotaped study allowed for more finely detailed analysis, including being able to link the table log data to the coded observation data. The naturalistic study is more closely representative of unobserved visitor behavior, certainly with respect to dwell time, and likely in other aspects as well. By including both types of visitor data, we can better extrapolate to actual exhibit impacts.

In addition to the evaluation studies of the DeepTree exhibit, a small add-on study focusing on the BAT exhibit was also conducted. One day of video/audio on the BAT exhibit (December 16, 2012: 3 hours, 52 minutes) was recorded, and one day of naturalistic observations (December 18, 2012: 6 hours, 3 minutes) was coded in real time. The observation coding protocol developed for the DeepTree exhibit was used without modification for the BAT exhibit observations. Although the protocol was not specifically developed for the BAT exhibit, the evaluators thought the BAT exhibit was similar enough to the DeepTree with respect to content and installation on the multi-touch table that the protocol would provide some interesting and relevant evaluation data. The BAT software was not designed to collect table log measures, so these were not available for the BAT studies. Surveys for a subset of study participants were administered on both observation data collection days. Table 1 provides a brief summary of the four evaluation studies conducted.

Table 1: Summary of exhibit studies and data collected

| Study | Exhibit focus | Method of data collection | Time | Data collected |
|---|---|---|---|---|
| DeepTree Video Study | DeepTree/ FloTree | Video/ Audio Recorded | 6 days, including one evening event: 16 hrs, 27 min | Event and Person data coded (N = 170)<br>Survey data (N=132)<br>Table log data collected and synchronized (N = 169) |
| DeepTree Naturalistic Study | DeepTree/ FloTree | Naturalistic | 3 days, including one museum “free day” and one evening event:<br>14 hrs, 31 min | Event and Person data coded (N = 326)<br>Survey data (N = 33)<br>Table log data collected, but unable to synchronize |
| BAT Video Study | BAT | Video/ Audio Recorded | 1 day: 3 hrs, 52 min | Event and Person data coded (N = 23)<br>Survey data (N= 19) |
| BAT Naturalistic Study | BAT | Naturalistic | 1 day, including one evening event: 6 hrs, 3 min | Event and Person data coded (N = 156)<br>Survey data (N = 12) |

## Data Processing and Coding

Although we recorded video data from two cameras, the two perspectives didn’t provide dramatically different information, so we used data from one camera as the primary video feed. As noted above, in addition to the audio from the video camera, there were four audio tracks recorded on two devices. Using Final Cut Pro (Apple, 2009) and PluralEyes (Red Giant, 2012) software, we synchronized and merged these audio tracks with the video (using the low quality sound track accompanying the video to synchronize). We then output a compressed Quicktime

Movie version of these files which, though still large (1.1-2.2 Gb per file) we could share with colleagues across the country.[8] Time stamps at the start of each video recording were noted to serve in the synchronization process.

As noted above, social interactions among people at the exhibit and with the table were coded from the video for the Video Studies (DeepTree and BAT); and directly in the field for the Naturalistic Studies (DeepTree and BAT).

Our intention had been to synchronize data for interactions among people – whether coded directly on the iPads on site at Cal Academy, or indirectly via the video – with data about activities done in the exhibit software, as recorded by the Table Log. However, when we went to make these links, we found that the times recorded directly on the iPads, the times recorded indirectly on the iPads from the videos, and the times recorded in the table log did not match. By looking carefully at the videos and the table logs, we were able to find points in time where we observed a specific action performed on the exhibit table that was clearly recorded in the table log. In this way, we were able to synchronize the Person, Event, and Table log data sources for the DeepTree Video study. We looked in a variety of ways for consistent patterns in the correction times needed to synchronize data in the video study, with the hopes that we could use those to extract rough table log information for the DeepTree Naturalistic study. Unfortunately, our best efforts in this direction yielded table log data that were clearly erroneous – with substantial times when no actions were occurring when we thought people should be at the table. Therefore, we decided we were unable to get usable table log measures for the naturalistic study. There was no table log data for the BAT study.

Survey data that were completed on paper forms were entered into a database by a TERC research assistant. Survey data were linked to other forms of data via sticker numbers in the video study. No such affirmative link was possible in the naturalistic study, and survey Ns were sufficiently small (10% or less) that it didn't seem useful to make rough links based on coordinating survey start times with exit times from observation data.

Thus, the only completely linked data set is from the DeepTree Video study.

## Analytic Methods

This evaluation was designed to gather a variety of information to understand how visitors worked with, responded to, interacted around, and learned from the exhibits. The aim of the data collection was to provide not only quantitative data about visitor behavior and how different variables such as group size relate to behavior, but also to capture rich, descriptive accounts of how visitors interacted with the exhibit and with each other around the exhibit. Our analyses comprised both exploratory investigation of the data as well as some hypothesis testing of specific questions.

Thus, our analyses include initial descriptive information, including demographic summary information about visitors, such as age, gender, education, and familiarity with touch technology, as well as visitor group size. Given our decision to allow visitors to come to and leave the exhibit as individuals in both the video and naturalistic studies, group size is not a trivial variable to define and obtain, and we describe the methods that we used in more detail below. We provide summary statistics on visitors' own ratings of their exhibit experiences with respect to interest, enjoyment, learning, and interaction with others. Next, we examine dwell time at the exhibits for each of the different studies, and analyze how dwell time varies by other

[8] To comply with confidentiality restrictions of our Institutional Review Board (IRB), data were stored on secure servers and were shared through password protected and secure file transfer protocols.

variables, including age category of visitor, gender, and group size. We further examine the nature of the groups through a graphic display of connectedness between individuals by study. Next we present a narrative description of groups in the DeepTree Video study, and observed characterization of groups by type/composition of group.

An analysis of coded event data is then presented with the frequency of different visitor behaviors quantitatively summarized by study, and within study, by group size. Another view of visitor behavior at the exhibit is presented through a summary of frequencies of a subset of the automatically logged table activities for the DeepTree Video study. We then describe our attempt to create a measure of table engagement through the use of confirmatory factor analysis of selected table activities and group interaction variables; an attempt that was informative, but left us only with time spent as the most reliable measure of engagement.

To capture what meaning visitors are taking away from their exhibit experience, we analyzed their responses from different survey questions, and also examined how these were related to aspects of their exhibit visit. First, we present a qualitative summary of visitors' descriptions of their understanding of what the exhibits were about, grouped by keywords and themes, by study and exhibit. Then, we examine how these visitors' open-ended responses demonstrate a range of levels of understanding, and provide a couple of selected illustrative examples.

Finally, we show the results of quantitative analyses we conducted to examine the relationships among variables of interest. We use hierarchical linear modeling (HLM) in these analyses because visitors' experiences are correlated with the group of people with whom they experienced the exhibit, and not using HLM would violate the independence assumptions of linear regression. Doing so, we examine whether group size and other variables are significantly associated with an individual's dwell time as a proxy for engagement. We then use HLM to test whether knowledge of evolution and common ancestry briefly measured at the end of the exhibit visit are each related to hypothesized variables of interest.[9]

# Findings

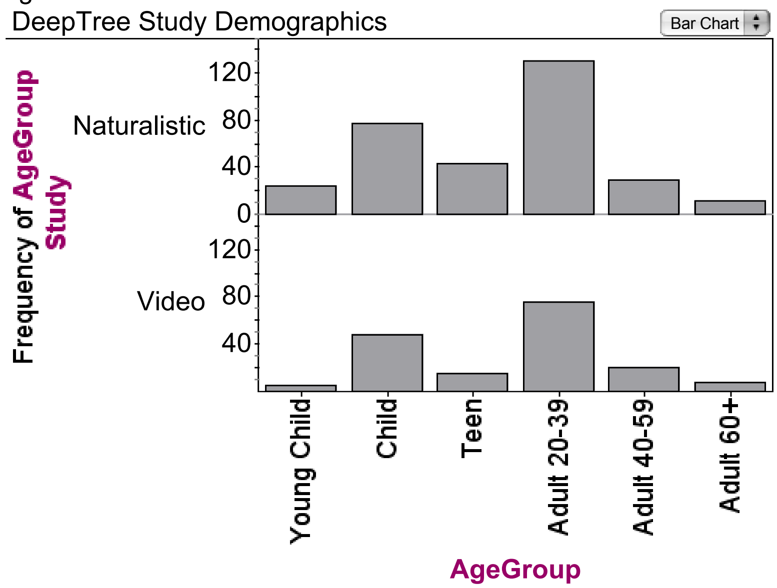
## *Description of Visitors by Study*

### DeepTree

The DeepTree Videotaped study included 170 participants who were categorized by age and gender by the evaluation observer. A subsample of these visitors (n =132; 78%) also completed surveys and provided age and gender information. A comparison of the estimated and reported ages indicated that estimated age group category were generally within one category of accuracy. Our observers tended to estimate younger ages for adults than they reported on the survey – i.e., observers "saw" more 20-39 year olds, and fewer 40-59 or 60+ year olds than participants themselves reported. This suggests that the observer-reported "ages" of adults in our naturalistic study may be slightly lower than is actually true. Ninety-five (95) participants were males (55.9%) and 75 participants were females (44.1%) in the DeepTree Video study.

[9] We used a wide range of software tools to conduct our analyses. These include SPSS (IBM, 2011), Excel (Microsoft, 2011), Fathom (Finzer, 2005), R (R Development Core Team, 2012), HLM (Raudenbush, Bryk, & Congdon, 2011), LISREL (Jöreskog & Sörbom, 2012), and NVivo (QSR International, 1999-2011).

Figure 6: DeepTree Age Distributions




Note: Based on observation data. Adult age classifications may be somewhat younger than self-reported ages. Video N = 170; Naturalistic N = 316.

The DeepTree Naturalistic Study involved a larger number of participants who were observed (n = 326) with 166 males (52.5%) and 147 females (46.5%), but a much smaller proportion who completed surveys (n = 33; 10%). Summary tables in the Appendices (Demographic Information, pp. A-3 to A-7) detail participants' gender and age information, both from estimates from the observation sample and from those who completed surveys. Figure 6 shows the distribution of observer coded ages for each of the studies. There are no statistically significant differences in these distributions ($\chi^2$ (df=5) = 7.56; $p$ = .18) suggesting the age distribution of our video study was representative of those who visited the exhibit in the naturalistic study.

Those completing surveys self-identified their race and ethnicity; observers did not attempt to identify race and ethnicity of participants by observation alone. Nearly three-quarters of participants in the DeepTree Video study (n = 96; 73%) identified themselves as white, 12% (n=16) identified as Asian American, with other races all less than 5% each. Nine (9) individuals (7%) identified as Hispanic or Latino/a. In the much smaller survey sample from the DeepTree Naturalistic study, roughly the same percentage of respondents indicated they were white (n = 21, 64%), Asian-American (n = 5; 15%) and Hispanic or Latino/a (n = 4; 12%). A breakdown of participant race and ethnicity can be found in the Appendices (Demographic Information, pp. A-3 to A-7).

About 84% of participants in the DeepTree Video study (n = 111) indicated that they spoke English at home, with over 25% of those (n = 29) speaking another language as well, including Spanish (n=13, 9.8% of total sample), Chinese (n = 8; 6.0%), and several languages that are each less than 5% of the total (French, Tagalog, Russian, Arabic, Italian, and Latin). Of the 16% (n = 21) who did not indicate that they spoke English at home, languages included French, Japanese, Chinese, Italian and German. Four (n = 4, 3%) participants did not indicate any language(s) they spoke at home.

Nearly all survey participants in the DeepTree Naturalistic study (n = 32, 97%) spoke English, including 15 (46%) who also spoke another language at home. The other languages identified included Spanish (n = 4, including the participant who did not speak English at home), and German (n = 3), as well as Italian, American Sign Language, Chinese, Tagalog, Arabic, Marathi, French, Russian, Polish, and Japanese. There may be some bias in the small survey sample for the naturalistic study, with fewer non-English speakers responding.

Individuals were also asked to identify their home state or country, if applicable. Of the 127 responses provided for the DeepTree Video Study, 48 (38%) individuals were from California and another 10 were from Mountain states (e.g. Colorado, Utah, Arizona, Nevada). Twelve (12) individuals were from the Northeast states (Massachusetts, Connecticut, New York, Pennsylvania), 10 were from the Central US, and 6 were from the South. Eleven (11) participants simply identified their location as the United States. Six (6) individuals were from Canada and 3 were from Mexico. Interestingly, 22 (17%) participants were from other non-North American countries (e.g. France, Argentina, England, the Philippines).

Of the 32 participants providing information about their home state or country for the DeepTree Naturalistic study, a much higher proportion, 66% (n = 21), were native to California. Another seven were US residents from other states. Several participants (n = 4, 12%) identified other countries of origin including Canada, Germany, and Sweden. Again, this small survey sample may not be representative.

All in all, Cal Academy seems to attract a racially, linguistically, and geographically diverse audience.

Participants were asked how much education they had completed. A table by age category for the DeepTree Video study is shown in the Appendices (Demographic Information, pp. A-3 to A-7). Adult visitors (aged 24+) to the exhibit are highly educated with 50% saying they have a 4-year college degree and an additional 32% saying they have a graduate degree. Of the sample that felt the question was applicable (n = 107), 14.4% of the total sample (n = 19) reported having a biology-related college degree.

Twelve of the 19 adult participants in the DeepTree Naturalistic study (63%) said they had a 4-year college degree or higher – slightly lower, but not statistically different from that in the Video study – and six participants (18%) indicated they had a biology-related degree. Reported education for visitors in the DeepTree Naturalistic study is shown in the Appendices.

The sample group of participants in the DeepTree Video study are also regular museum-goers, reporting an average (mean) of 3.3 museum visits in the past year (*SD* = 4.25; *median* = 2; *range* 0-15). Just over one-third (34.9%, n = 46) of participants reported having used a touch table before their visit to the museum on the day of the study. To assess the visitors' familiarity with other mobile/ touch screen devices, they were asked to rate how much they have used iPads, smart phones, and other touch devices on a scale from 1 (not at all) to 5 (a lot). On average, participant ratings were quite high, at 4.27 (*SD* = 1.00).

For the DeepTree Naturalistic study, the number of reported museum visits in the last year was highly variable (*range* 0-200; *median*=2), but over half (n = 18, 55%) reported one to three visits. About 36% (n = 12) indicated that they had used a touch table prior to their experience with this exhibit, and most participants reported familiarity with touch devices, with a mean use rating of 4.16 (*SD* = 1.31) on a scale from 1 (not at all) to 5 (a lot) – again quite high.

While the "modal" visitor to the DeepTree exhibit, whether in the Video or Naturalistic study, is a white, college-educated, English-speaking, 20-something, male, from California, with museum and technology experience, there is substantial variation as noted above in all of these categories. There are no statistically significant demographic differences between the Video and Naturalistic studies.

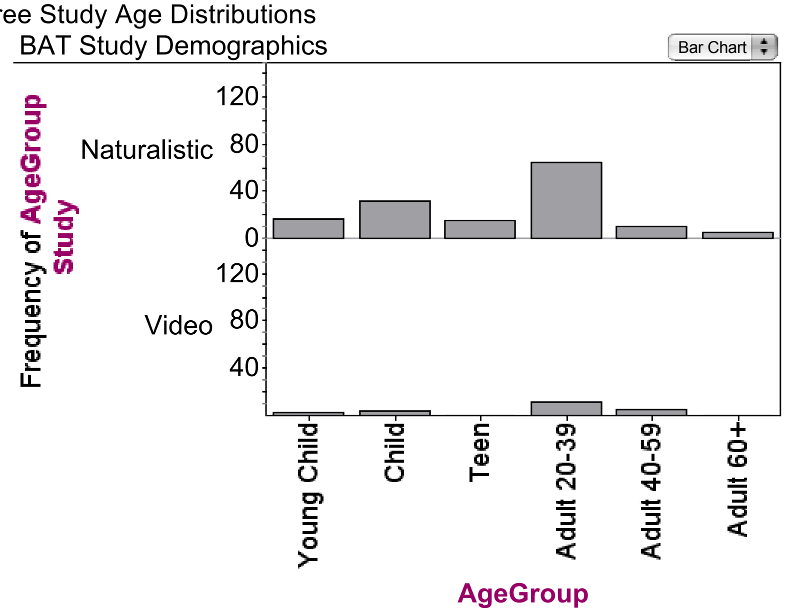
## Build-A-Tree

The two BAT studies were smaller in scale. Of the videotaped participants (n = 23), 52% were male and 39% were female (observers couldn't identify gender for 2 participants), and 19 (83%) completed surveys. A total of 156 visitors participated in the BAT naturalistic study, 51% male and 49% female, but only 12 of these (8%) completed surveys. A graph of the age distribution of the BAT studies is below, and tables of these and other demographic data appear in the Appendices (pp. A-6 to A-10).

Because of the small samples completing surveys in these two studies, caution should be used in making any inferences from these data. For the BAT Video study, 42% of survey respondents said they were white and 37% said they were Asian-American, with 42% also saying they were Hispanic or Latino/a. Three-quarters of BAT Naturalistic survey respondents (75%) said they were white with all other Ns too small to report.

In the BAT video study, 17 (89%) of the participants spoke English, five spoke Spanish, one spoke Chinese, and 4 others reported speaking Arabic, German (non-English speaking), Kannada (non-English speaking), and Tamil (both south Indian languages). Twelve of these individuals were native to California and another five mentioned residence in other parts of the US. There was also one participant from Canada and another from Germany. Among those in the BAT naturalistic study, 11 spoke English, two spoke Spanish, one spoke Chinese and one spoke Serbian (non-English speaking). Six of these individuals were living in California, three were from Australia, one from Serbia and one from Singapore.

Figure 7: Build-A-Tree Study Age Distributions




Note: Based on observation data. Adult age classifications may be somewhat younger than self-reported ages. Video N = 23; Naturalistic N = 156.

Five of the 8 adult respondents (63%) in the BAT video study said they had 4-year college degrees or higher, and eight individuals (42%) indicated they had a degree that was biology related. For the BAT naturalistic study, five of the seven adult participants (71%) reported a 4-year college degree or higher, and two participants reported having degrees in a biology-related field. Details of age and education data for both studies appear in the Appendices (pp. A-6 to A-10).

The number of museum visits in the past year among the BAT video study participants was variable, ranging from 0-25. The median number of visits in the past year was 2. Thirty-seven percent (37%) of participants in this group had used a touch table before their visit on the day of data collection, comparable to the proportions in the DeepTree studies. When asked on a scale of 1-5 how much they've used touch devices prior to their visit, the average rating was 4.58 (*SD* = .90).

The number of museum visits in the past year among the BAT naturalistic study participants was also quite variable, ranging from 0-60. The median number of visits in the past year was 2.5. Fifty percent (n=6) of participants in this group had used a touch table before their visit on the day of data collection. When asked on a scale of 1-5 how much they had used touch devices prior to their visit, the average rating was 3.83 (*SD* = .94).

### *Exhibit Experience*

Visitors to both the DeepTree and BAT exhibits generally found their experience interesting and enjoyable, and said they learned from their experience and from working with others at the exhibits.

Survey participants were asked to rate their perceptions of the exhibit experience. They used a 5-point scale for all these ratings, with slightly different meaning depending on the question — i.e., 1 = not interesting/did not like/learned nothing; 5 = very interesting/ liked a lot/learned a lot, etc. Participants answered questions about their interest, enjoyment, and learning at the exhibit, and how working with others at the table made a difference in these experiences. Average ratings for the sample can be found in Table 2. Participants reported high levels of interest and enjoyment with the exhibit (average ratings > 4), and moderate levels of learning (average rating 3.37).

Table 2: Self-reported ratings of exhibit experience

| Question | DT Video | DT Naturalistic | BAT Video | BAT Naturalistic |
|---|---|---|---|---|
| How interesting was the touch table? | 4.13 (0.94) | 4.24 (0.83) | 4.28 (0.46) | 4.33 (0.65) |
| How much did you like using the touch table? | 4.10 (0.94) | 4.15 (0.94) | 4.44 (0.62) | 4.50 (0.67) |
| How much did you learn at the touch table? | 3.37 (1.16) | 3.27 (0.94) | 3.89 (0.94) | 3.67 (0.65) |
| How did others at the table affect your learning? | 3.28 (1.27) | 3.09 (0.95) | 4.29 (0.83) | 3.67 (1.12) |
| How much did you like working with others at the table? | 3.54 (1.23) | 3.46 (0.98) | 4.50 (0.52) | 4.20 (1.23) |
| Would you tell a friend to visit the table? | 3.98 (1.22) | 4.24 (1.17) | 4.16 (1.07) | 4.25 (0.87) |

Note: All ratings on a 5 point scale. Ratings are Mean (SD).
DeepTree Video Study, N = 132. DeepTree Naturalistic Study, N = 33.
BAT Video Study, N = 19. BAT Naturalistic Study, N = 12.
Ns slightly lower on questions about working with others at the table.

Since the social nature of learning at the exhibit was also important, we asked participants to use 5 point scales to rate the extent that having others at the table affected their learning (1 = much harder, 3 = did not affect, 5 = much easier) and enjoyment (1 = did not like, 3 = some, 5 = liked a lot). Both questions yielded neutral to slightly positive responses, with more variability than the simple enjoyment and learning questions. Finally, respondents used a 5-point scale to say whether they would recommend the table to a friend (1 = No, 3 = Maybe, 5 = Yes), averaging a clearly positive 3.98 ($SD = 1.21$). Ratings of exhibit experience in the DeepTree Naturalistic study produced similarly positive results (see Appendix, Survey Results, p. A-41 for more detailed statistics).

When posed the same questions in the BAT Video Study, visitor ratings were similar, though somewhat higher on average for the social interaction questions. That is, when using the BAT software in the video study, people seemed to feel that working with others at the table made it easier and increased their enjoyment more than when working with DeepTree in either the video or naturalistic conditions.

When posed the same set of questions in the BAT Naturalistic study, visitors gave similar high ratings for their interest, enjoyment and learning with the exhibit software, and ratings of how having others at the table affected their learning and enjoyment were again more in line with ratings from the DeepTree video and naturalistic studies.

Thus, it seems that when exploring BAT, visitors felt that working with others had a more positive impact on their enjoyment and sometimes their learning, than when working with DeepTree. It's not clear what this difference means — perhaps navigating the simpler BAT software with a more familiar group led to these more uniformly positive results.

## *Dwell Time*

### DeepTree Exhibit

As one measure of engagement with the exhibit content, we conducted a variety of analyses on the amount of time that visitors spent at the table. Distributions of dwell time were positively skewed, with many people spending a relatively small amount of time and a few people spending a long time. To describe typical time at the table with a skewed distribution, one can either report the median (middle value), which is not influenced by extreme values; or one can transform the data to a different scale where they're more normally distributed (in this case, a log transformation works). We do both, reporting statistics calculated on the log-transformed data, but reported in natural units for clarity here; and median values with interquartile ranges in the Appendix (Dwell Time Details*,* p. A-17). We begin by reporting descriptive statistics, not controlling for group membership or other group characteristics. Later, after introducing the full set of measures and analyses, we report on dwell time controlling for size of group and other characteristics.

Exhibit dwell times differed by study, with participants in the DeepTree Video study spending longer, on average, than those in the DeepTree Naturalistic study; and those in the BAT Video study spending longer on average than those in the BAT Naturalistic study. Given the consent process involved in the videotaped study, we hypothesize that participants in the video study had a greater commitment to trying exhibit features than those in the naturalistic context and may have felt some sense of responsibility for giving us good data.

Table 3: Time Spent at Table by Age Category

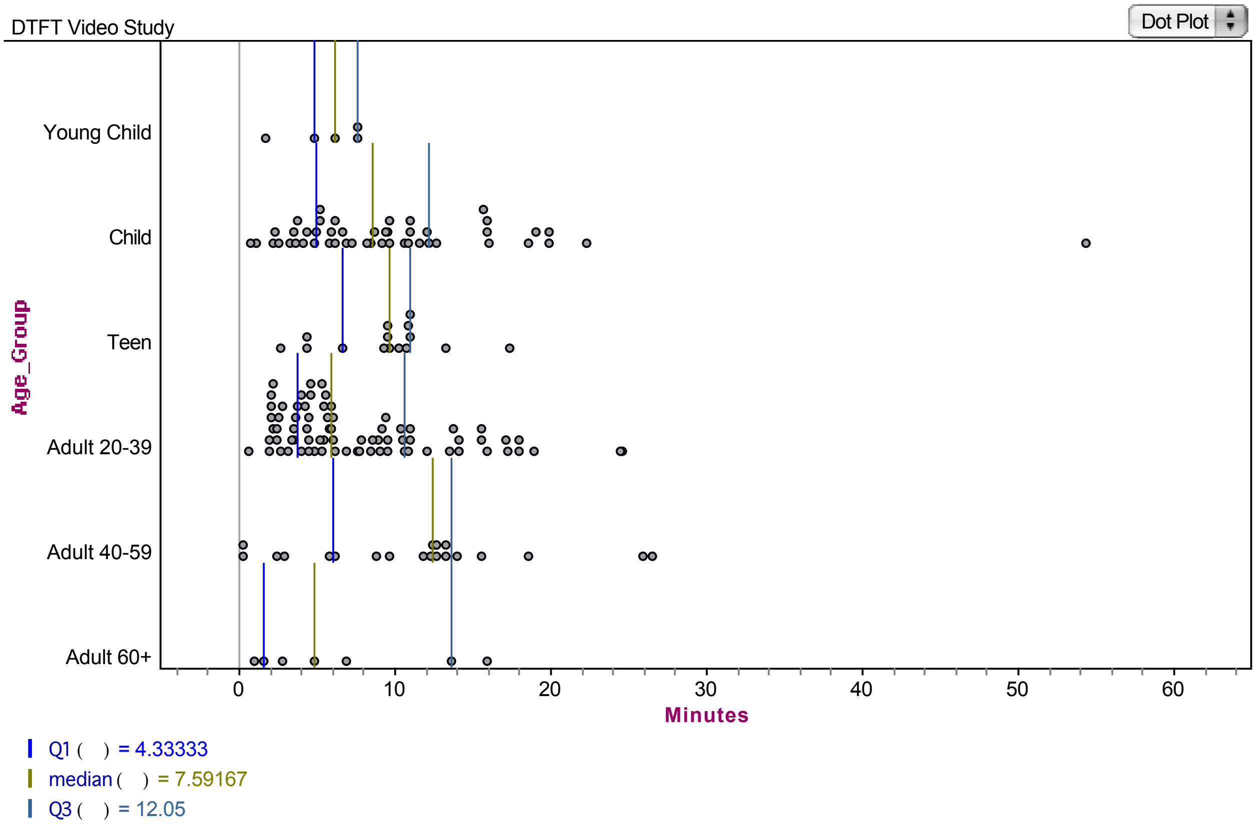
| Age Category | DeepTree Video | DeepTree Naturalistic | BAT Video | BAT Naturalistic |
|---|---|---|---|---|
| Young Child | 296 s (4.93 min) | 44 s (0.73 min) | 527 s (8.78 min) | 81 s (1.35 min) |
| Child | 446 s (7.44 min) | 55 s (0.92 min) | 1215 s (20.25 min) | 143 s (2.38 min) |
| Teen | 510 s (8.51 min) | 63 s (1.05 min) | – | 118 s (1.97 min) |
| Age 20-39 | 368 s (6.13 min) | 90 s (1.50 min) | 385 s (6.41 min) | 112 s (1.86 min) |
| Age 40-59 | 440 s (7.33 min) | 105 s (1.75 min) | 616 s (10.27 min) | 138 s (2.30 min) |
| Age 60+ | 262 s (4.37 min) | 70 s (1.17 min) | – | 44 s (0.73 min) |
| Unknown | | | 403 s (6.72 min) | 35 s (0.58 min) |
| **Total** | **400 s (6.67 min)** | **69 s (1.14 min)** | **502 s (8.36 min)** | **102 s (1.71 min)** |

Note: Table shows mean log-transformed dwell time, in natural units; Seconds (Minutes).
DeepTree Video Study, N = 169. DeepTree Naturalistic Study, N = 326.
BAT Video Study, N = 23. BAT Naturalistic Study, N = 156.

*By age, gender, and visit day*

In the DeepTree Video study, based on log-transformed data, average time at the exhibit was 6.67 minutes (400 seconds), median time was 7.58 minutes (455 seconds) with the middle 50% of visitors spending between 4.3 and 11.9 minutes at the exhibit. (This means that 25% of visitors spent less than 4.3 minutes; and 25% spent more than 11.9 minutes.) The minimum amount of time spent by a person was 15 seconds, the maximum was 54.3 minutes (3259 seconds). These are long average dwell times for a museum exhibit and suggest that, at least in the video study, people were quite engaged with the DeepTree software. Because untransformed mean dwell times are often reported in exhibit evaluation studies, we will also report these – in this case, 530 seconds or 8.83 minutes – but this is clearly an overestimate. No gender differences were found. Dwell time by age category can be found in Table 3, with a plot of these data shown in Figure 8. There were no statistically significant differences in dwell time by age category ($F$ (5, 163) = 1.122; $p$ = .351).

Figure 8: Graph of Time Spent at Table by Age Category for DeepTree Video Study.

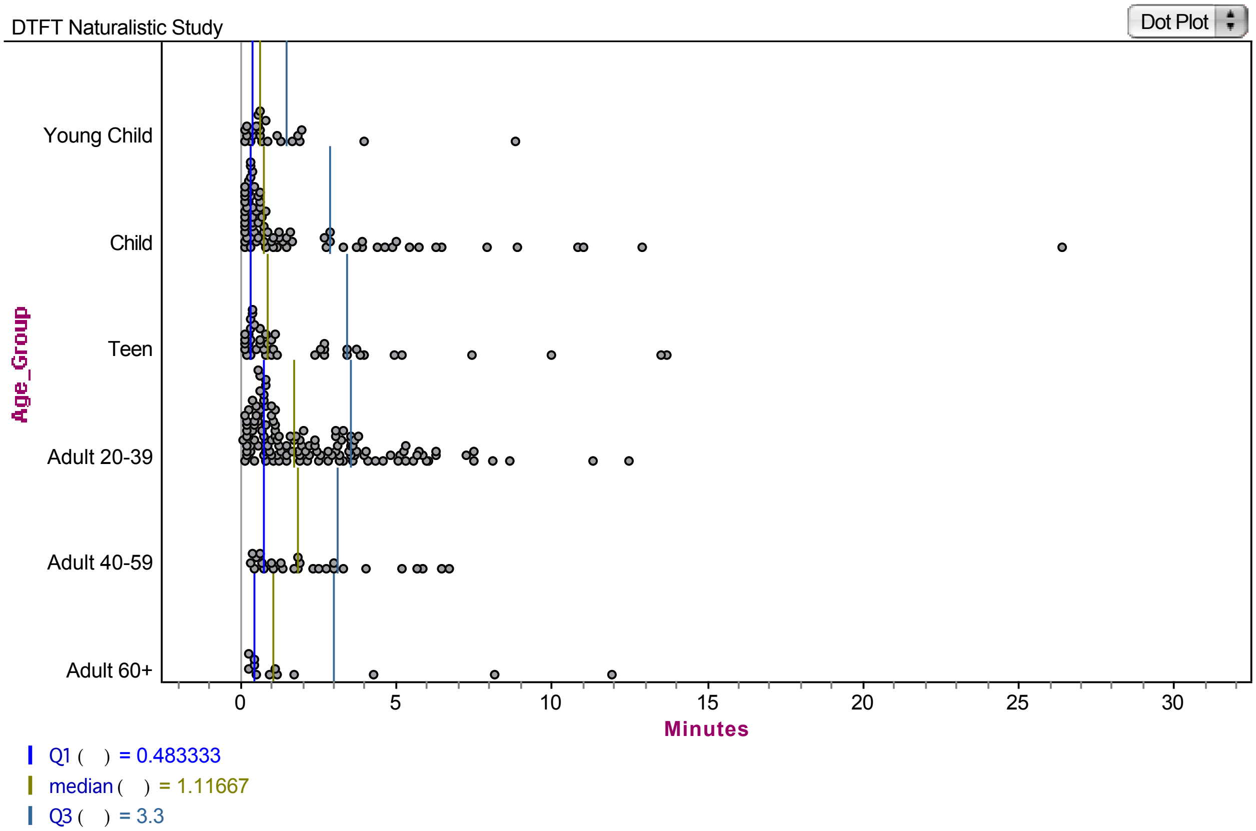



**Notes**: This graph omits one child who spent 54.32 minutes (3259 seconds) at the exhibit. Plotted lines represent the 25th percentile, 50th percentile (Median), and 75th percentile of time spent for each age group.

As a component of the survey, participants were asked to self-report how many minutes they spent at the touch table. For individuals with both estimates of time spent from the video observation study as well as the survey, participant self-reports and estimates of time spent from the video were significantly correlated, albeit a small to moderate association ($r = .23$, $p = .008$).

In the DeepTree Naturalistic Study, based on log-transformed data, individuals spent, on average 69 seconds (1.15 minutes) at the exhibit (mean of untransformed data is 137 seconds (2.28 minutes)). The median time was 67 seconds (1.12 min) with the middle 50% of visitors spending between 26 seconds and 194 seconds (3.23 minutes). The minimum amount of time spent by a person was 3 seconds, the maximum was 1584 seconds (26.4 minutes). There was no difference in time spent by gender, but there was a statistically significant difference in time spent (log-seconds) by age category overall ($F$ (5, 310) = 3.36; $p = .006$). Data were collected over a period of three diverse days; analysis indicated no significant differences in dwell time across the different days. Dwell time by age category can be found in Table 3, and a plot of these data can be found in Figure 9.

Figure 9: Graph of Time Spent at Table by Age Category for DeepTree Naturalistic Study.




**Note**: Plotted lines represent the 25th percentile, 50th percentile (Median), and 75th percentile of time spent for each age group.

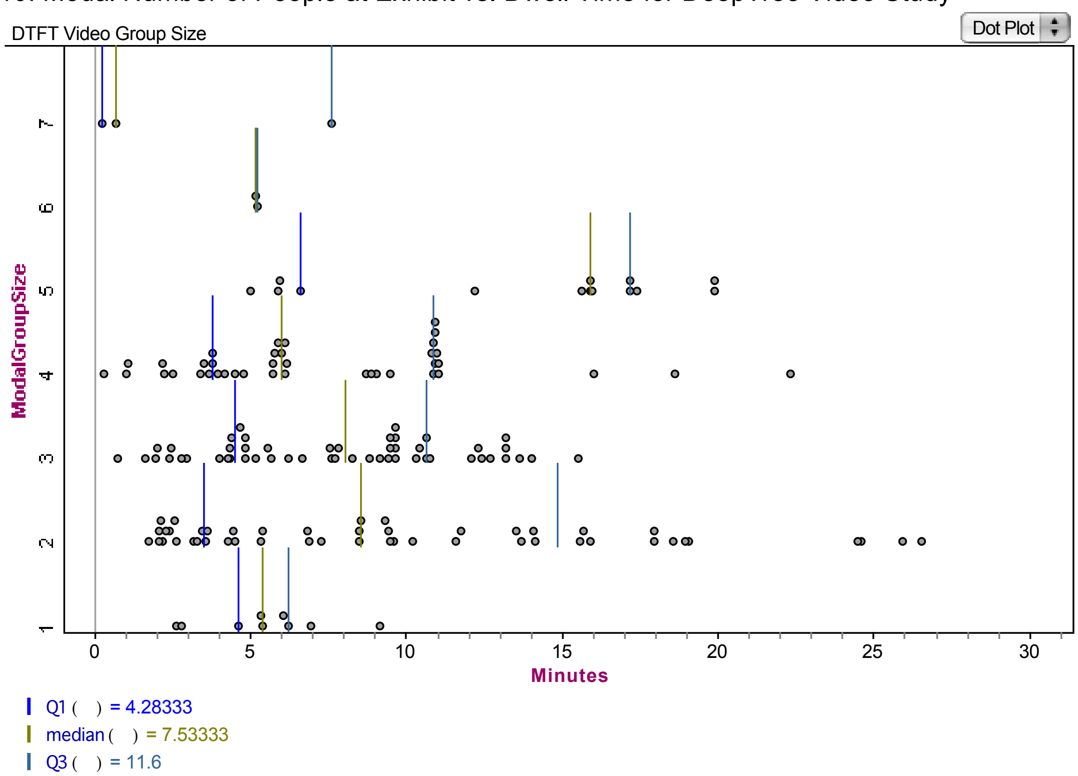
*By group size*

In the naturalistic museum context, people come to the exhibit table and leave in a wide variety of patterns and ways – sometimes as individuals, sometimes in groups, sometimes splitting up and coming back together, and so forth. This means that the "group" gathered around the exhibit at any one time may be related to one another, may know each other well, or may be strangers. To gather evidence about people's experiences in a way that modeled this fluid behavior both in the naturalistic and video studies (as long as they consented), we allowed visitors to come and go at will rather than control access to the table to particular types of groups and configurations. This meant that for any individual, the "group" with whom they shared experiences at the exhibit table varied over time. But it was important for our analyses to have some measure of group size for each individual – to be able to describe the typical number and configuration of people around the table during the time that each person was there. It was also important for our hierarchical models that people were identified with a single group who could be said to share most of their experiences together.

We thought about how to capture "group" in this fluid situation in a couple of ways. If someone arrived as others were just leaving, it seemed reasonable that their interactions wouldn't influence one another very much. What if they overlapped 25% of the time? 35% of the time? Clearly, if they overlapped for at least half of the time, they could be considered part of one another's group. We decided to set the cut off at 40% – that is, we calculated the **modal group size** for each individual by counting the number of others who were present at the exhibit with them for at least 40% of their time at the table (using Event coded 20-second intervals as a proxy). For some people who stayed at the exhibit while others came and went, this meant that there were very few people who were part of their "group," even if a large number of others had been at the table for some portion of their time.

At the same time, with this method, groups were not entirely exclusive of one another – someone could be part of two other people's modal "group" even if those others were not in a group together. For our hierarchical analyses, we needed to develop a way of defining **exclusive groups** and minimizing this type of overlap. Thus, we looked for groups of people who *all* overlapped one another half the time or more. In a few cases, we also grouped people where A only overlapped B for 40% of B's time, but B overlapped A for at least 60% of A's time. For the DeepTree Video study, we identify 73 exclusive groups. This method wasn't perfect – there were still two overlapping "exclusive groups", but we assigned individuals who could have been in more than one group to the groups with whom they spent more of their time. These exclusive groups are used in hierarchical regression models when we need to account for the non-independence of the data (see Nested Models for Time Spent, page 54, below.)

Figure 10: Modal Number of People at Exhibit vs. Dwell Time for DeepTree Video Study




**Notes**: This graph omits one child in a group of 3 who spent 54.32 minutes (3259 seconds) at the exhibit. Plotted lines represent the 25th percentile, 50th percentile (Median), and 75th percentile of time spent for each group size.

In the DeepTree Video study, the average modal group size was 3.13 people (*SD* = 1.22), *range* = 1 to 7 people, median group size of 3. Table 4 and Appendix, Dwell Time Details, pp. A-17 to A-20 show the average exhibit dwell time by modal number of people at the table during an individual's experience, and Figure 10 graphs these.

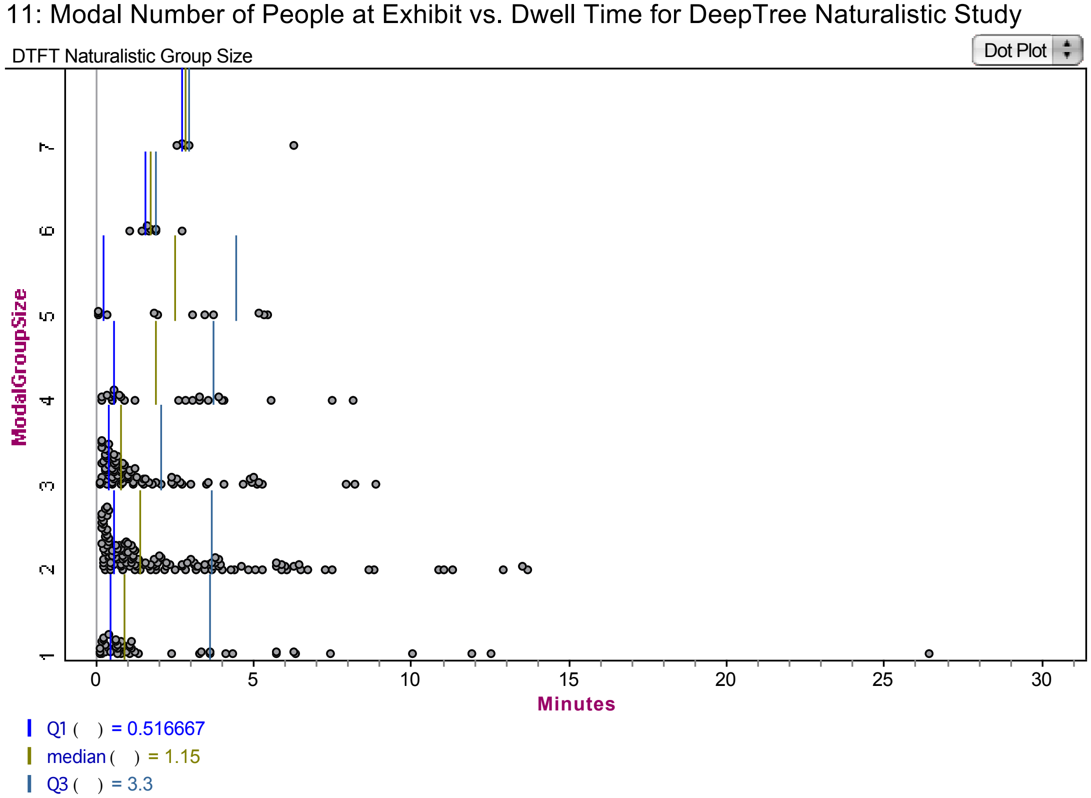
Table 4: Modal Number of People at Exhibit vs. Dwell Time

| Group Size | DeepTree Video | DeepTree Naturalistic | BAT Video | BAT Naturalistic |
|---|---|---|---|---|
| 1 person | 295 s (4.91 min) | 68 s (1.13 min) | 447 s (7.46 min) | 72 s (1.20 min) |
| 2 people | 481 s (8.02 min) | 85 s (1.42 min) | 646 s (10.76 min) | 110 s (1.83 min) |
| 3 people | 401 s (6.68 min) | 55 s (0.91 min) | 264 s (4.39 min) | 109 s (1.81 min) |
| 4 people | 336 s (5.60 min) | 85 s (1.41 min) | 374 s (6.23 min) | 100 s (1.67 min) |
| 5 people | 826 s (13.77 min) | 64 s (1.06 min) | | 171 s (2.84 min) |
| 6 people | 110 s (1.84 min) | 103 s (1.71 min) | | 126 s (2.10 min) |
| 7 people | 25 s (0.42 min) | 195 s (3.24 min) | | |
| **Total** | **400 s (6.67 min)** | **69 s (1.14 min)** | **502 s (8.36 min)** | **102 s (1.71 min)** |

Note: Table shows log-transformed mean dwell time, in natural units; Seconds (Minutes).
DeepTree Video Study, N = 169. DeepTree Naturalistic Study, N = 326.
BAT Video Study, N = 23. BAT Naturalistic Study, N = 156.
There is 1 person in the BAT Naturalistic study for whom we do not have group size information.

Notice that average time at the exhibit is smaller for a person alone at the table compared to a group of two or more people. Notice also that average time at the exhibit decreases slightly from groups of 2 to groups of 3 to groups of 4, but then goes back up dramatically for groups of 5. This suggests that social interaction at the table may help people engage with the exhibit, to a point. However, it's not clear whether these patterns of relationship between group size and

Figure 11: Modal Number of People at Exhibit vs. Dwell Time for DeepTree Naturalistic Study




**Notes**: Plotted lines represent the 25th percentile, 50th percentile (Median), and 75th percentile of time spent for each group size.

dwell time are robust. We will examine them further, accounting for the shared group experience in, the section on *Nested Models*, below.

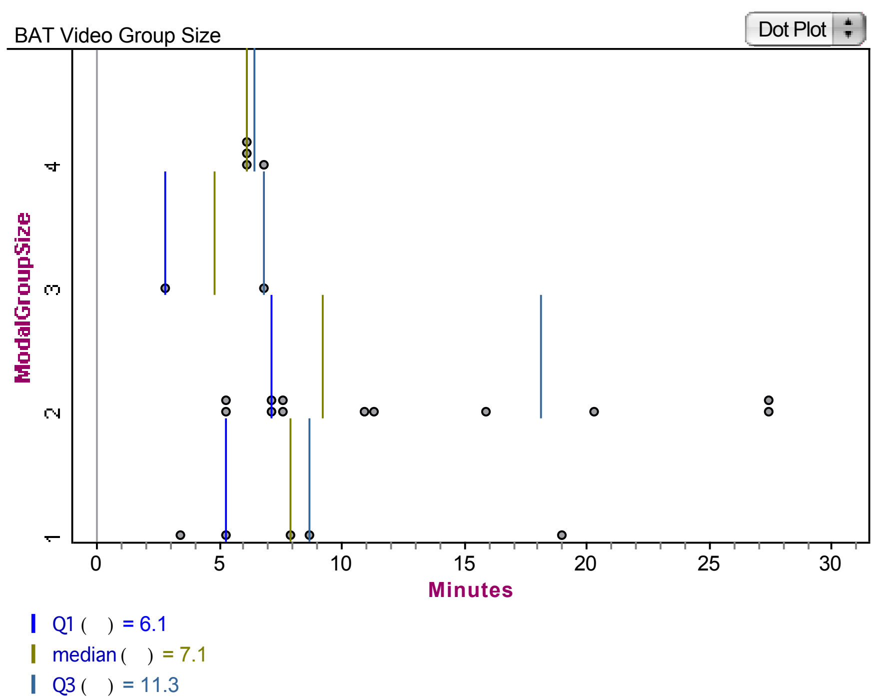
In the DeepTree Naturalistic study, the average group size was slightly smaller than in the video study – 2.56 people (*SD* = 1.25), range = 1 to 7 people, median group size was 2. Table 4 and Figure 11 show the average exhibit dwell time in the naturalistic study based on the average number of people at the table during the individual's experience. Here we see more erratic patterns of relationship between group size and dwell time. That is, the patterns we saw with the Video study don't seem to be replicated in the Naturalistic study. Again, we will explore these further in the section on *Nested Models*, below.

## BAT Exhibit

We also gathered dwell time data for the BAT exhibit, in both the video and naturalistic studies, and examined dwell time with respect to age category, and group size. As with the DeepTree data, we used log-transformed data to calculate measures of central tendency, and then transformed back to natural units for reporting. Similar to the findings with the DeepTree exhibit, participants in the BAT video study tended to have longer dwell times than in the BAT naturalistic study.

Across the 23 BAT Video study participants, based on log-transformed data, individuals spent on average 502 seconds (8.36 minutes) at the exhibit (mean of untransformed data is 10.08 minutes) (see Table 4 and Appendix, Dwell Time Details, pp. A-17 to A-20 for details). The median time was 424 seconds (7.07 minutes) with a minimum time of 2.83 minutes, a maximum of 27.43 minutes, and the middle 50% of visitors spending between 6.05 and 11.32 minutes.

Figure 12: Modal Number of People at Exhibit vs. Dwell Time for BAT Video Study

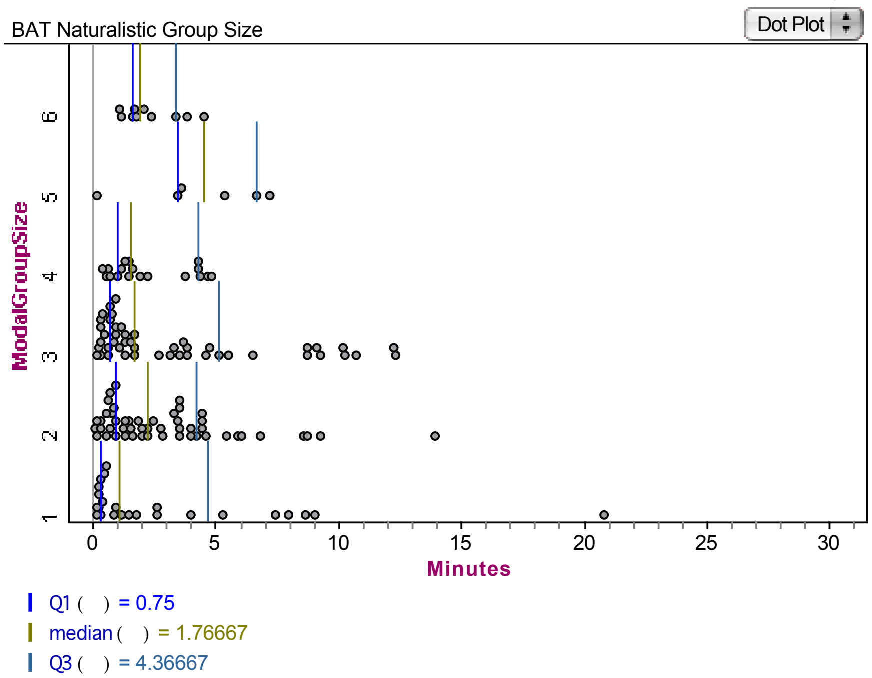



In the small BAT Video study, the average group size was 2.22 people (*SD* = 0.99), *range* = 1 to 4 people, median group size was 2. Table 4 and Figure 12 show the average exhibit dwell time based on the average number of people at the table during the individual's experience.

There is no statistically significant difference in dwell time by group size (*F* (3, 19) = 2.132, *p* = .130).

Across the 156 BAT Naturalistic study participants, based on log-transformed data, individuals spent on average 102 seconds (1.71 minutes) at the exhibit (mean of untransformed data is 3.08 minutes) (see Table 4 and Appendix, Dwell Time Details, pp. A-17 to A-20 for further details). The median time was 105 seconds (1.76 minutes) with a minimum time of 3 seconds, a maximum of 1249 seconds (20.82 minutes), and the middle 50% of visitors spending between 44 seconds and 4.35 minutes. There is no statistically significant difference in dwell time by age category (*F* (5, 138) = 1.382, *p* = .235).

Figure 13: Modal Number of People at Exhibit vs. Dwell Time for BAT Naturalistic Study




In the BAT Naturalistic study, the average group size was 2.75 people (*SD* = 1.32), *range* = 1 to 6 people, median group size was 3. Table 4 and In the small BAT Video study, the average group size was 2.22 people (*SD* = 0.99),
*range* = 1 to 4 people, median group size was 2. Table 4 and Figure 12 show the average exhibit dwell time based on the average number of people at the table during the individual's experience. There is no statistically significant difference in dwell time by group size (*F* (3, 19) = 2.132, *p* = .130).

Across the 156 BAT Naturalistic study participants, based on log-transformed data, individuals spent on average 102 seconds (1.71 minutes) at the exhibit (mean of untransformed data is 3.08 minutes) (see Table 4 and Appendix, Dwell Time Details, pp. A-17 to A-20 for further details). The median time was 105 seconds (1.76 minutes) with a minimum time of 3 seconds, a maximum of 1249 seconds (20.82 minutes), and the middle 50% of visitors spending between 44 seconds and 4.35 minutes. There is no statistically significant difference in dwell time by age category (*F* (5, 138) = 1.382, *p* = .235).

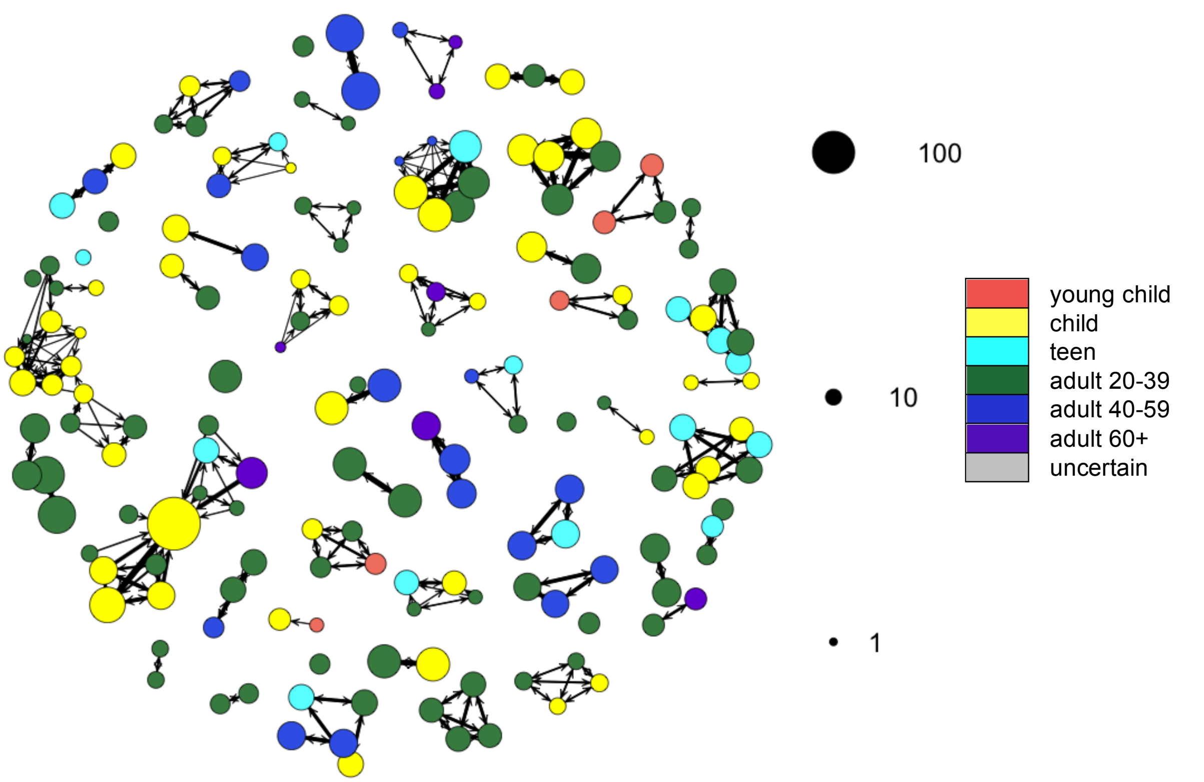
Figure 13 show the average exhibit dwell time based on the average number of people at the table during the individual's experience. There is no statistically significant difference in dwell time by group size ($F$ (5, 149) = .745, $p$ = .591).

## *Group Configurations and Interaction Patterns*

### Groupings

Graphic displays of the "groups" that formed in the different studies provide an image of who was with whom around the exhibit table. Data for these displays come from matrices representing which visitors were present at which coded 20-second event intervals. We multiplied this matrix by its transpose — shifting from a two-mode network of actors and 20-second events to a one-mode network of actors connected by common events (Wasserman & Faust, 1994) — in the process counting the number of 20-second intervals that each pair of visitors shared, and then displayed these using the R social network analysis module (Butts, 2013; R Development Core Team, 2012). In the displays below, color indicates age category (roughly rainbow order from hot colors for the youngest participants to cool colors for the oldest: salmon = young child; yellow = child; aqua = teen; green = adult 20-39; blue = adult 40-59; purple = adult 60+; grey = uncertain age). Area of dots is proportional to the time the person spent at the exhibit, and is consistent across the several displays — the black legend dots show a scale of 1, 10 and 100 events respectively. Width of the line connecting dots is proportional to the time people overlapped at the exhibit (thicker is more).

Figure 14: DeepTree Video Study Group Connections



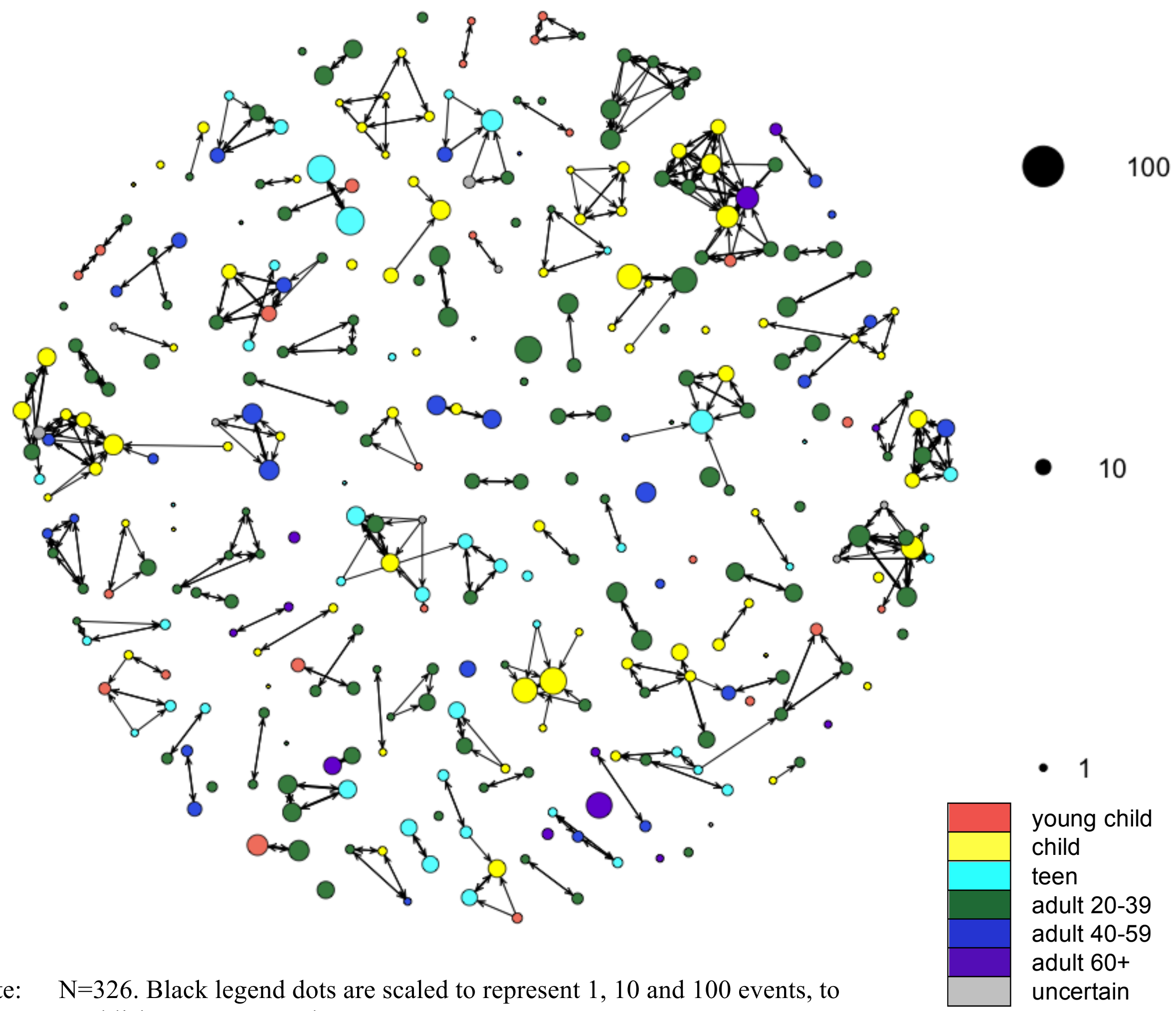

Note: N=169. Black legend dots are scaled to represent 1, 10 and 100 events, to establish a common scale.

We have filtered out connections that were less than 40% of both people's time at the exhibit table, so these graphs represent connections that contributed to modal group size.

Arrowheads point towards people with whom a person shared 40% or more of their time. Lines with two arrowheads are mutual connections, identifying members of "exclusive groups"; those with one arrowhead only "count" in the modal group size for the source of the arrow.

These displays show that those in the Video study stayed longer, had fewer singletons, and somewhat less complex patterns of overlap than those in the Naturalistic study. One statistical measure of this increased complexity in the Naturalistic study compared to the Video study is the average "betweenness centrality" – the extent to which "connections" between people pass through other people. For the Naturalistic study, average betweenness centrality is 2.39; for the Video study, just 1.03 suggesting that in the video study, people were more directly connected to one another rather than being "connected" only because they each shared time with a third person.

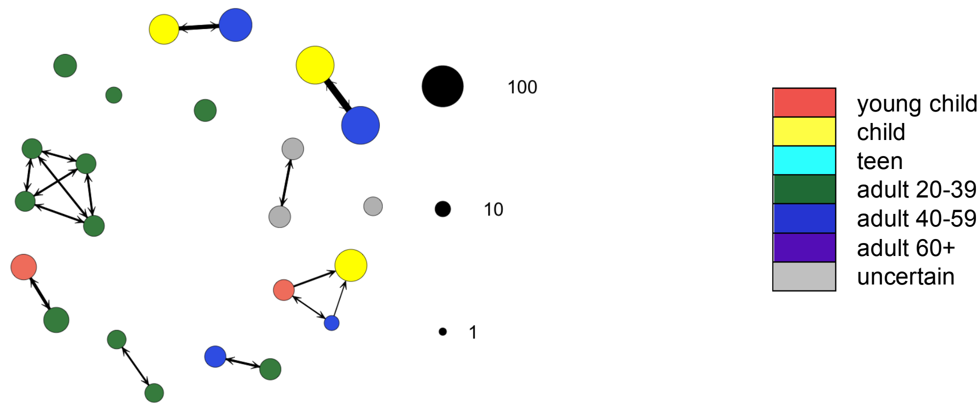
Figure 15: DeepTree Naturalistic Study Group Connections




Note: N=326. Black legend dots are scaled to represent 1, 10 and 100 events, to establish a common scale.

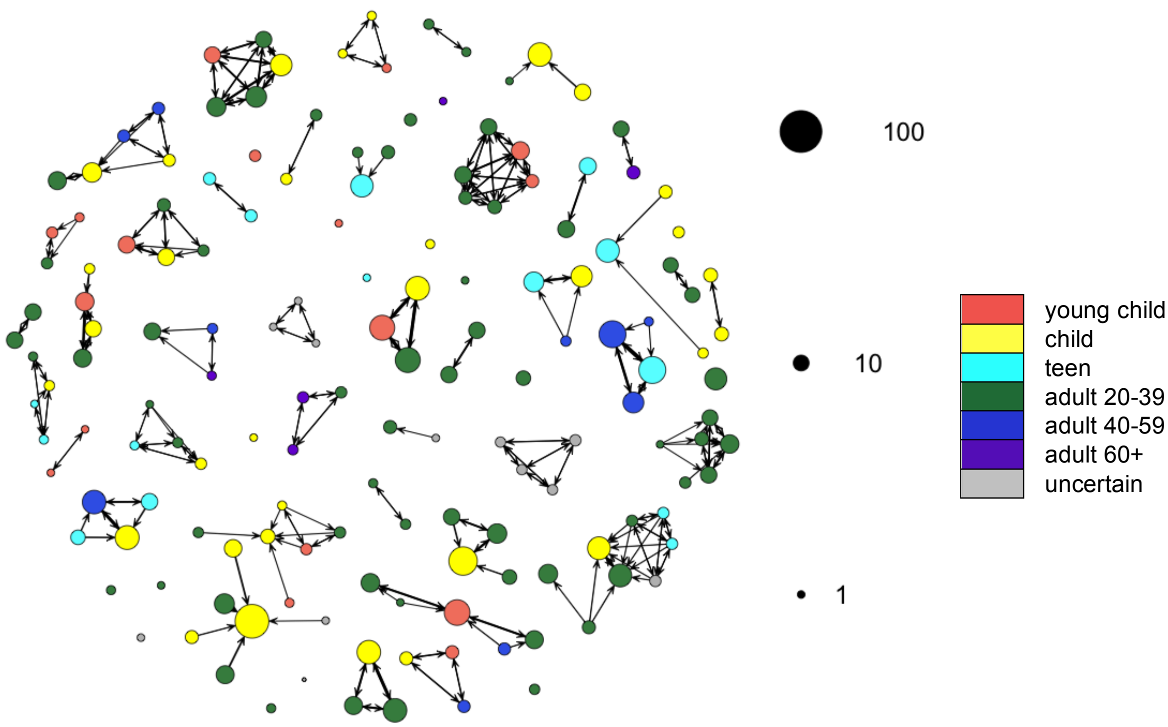
We conducted a similar analysis for groups around the BAT software exhibit, using the same conventions and scale. Again, those in the Video study tended to stay longer and had less complex connections. In fact, mean betweenness centrality for the BAT video study is zero (0) because people always had direct connections to others with whom they worked. For the BAT naturalistic study, mean betweenness centrality is 2.13, comparable to the DeepTree naturalistic study.

Figure 16: BAT Video Study Group Connections




Note: N= 23. Black legend dots are scaled to represent 1, 10 and 100 events, to establish a common scale.

Figure 17: BAT Naturalistic Study Group Connections




Note: N=156. Black legend dots are scaled to represent 1, 10 and 100 events, to establish a common scale.

## Group Composition

Several evaluators went through the videotapes from the DeepTree Video study to qualitatively characterize interaction patterns. Based on qualitative observations of group interactions, with groupings identified as individuals who appear to arrive and leave together, approximately 68 distinct visitor groups were recorded at the table. This is comparable but not exactly the same as the 73 groups identified from the more quantitative analysis above.

Of these, about 30 were "family" groupings of mixed adults and children. Eleven (11) were adult-child pairs; eight (8) were three person groupings and 11 were four or more people in the group. Some of these were single adults with two or more children; sometimes there were two adults and two more children (for two of the adult-child groups, the adults left after a little while at the table and let the children work by themselves). Two-thirds of the family groupings had just one adult present. Dwell time ranged from ~2:20-20:10, median time 9:20.

Nine were child-only groups, about half of the children were alone, half with another child. There also was what appeared to be a group of 6 boys in school uniforms who interacted with the table, supervised by a single adult who was not at the table very long. The children in these groups mostly appeared to range from about 7 to 12, with a few teens also present in the larger family groups.

Twenty-nine (29) groups were adult only. Of these, ten were single adults, 12 were in pairs, and 7 were in groups of three or larger. Some of these were groups of 20-something adults; others had older adults as well, and these behaved somewhat differently.

*Overlapping of Groups*

Because this was a consented videotaped study, with the video area cordoned off, there was not as much overlap between groups as was observed in the naturalistic study. However, there was some overlap, and a few examples are described here:

> *In one case, a boy approximately 8 or 9 years old arrives with his grandpa while a teen boy is at the table. The teen is looking at the DNA trait information. The younger boy enlarges some images from the canopy and then moves over to select a species from the reel. The teen observes for a while longer, watching what happens as the younger boy selects different choices at the table. During this time, a young adult couple enter the observation area. They watch for awhile but eventually leave without ever touching the table, which is essentially monopolized by the boy. The teen boy leaves after a bit as well. Grandpa comes and goes a few times during the 55 minutes that this boy is at the table.*
>
> *A 20-something woman comes to the table and also watches for awhile before leaving without touching the table. Then a 20-something man arrives, while grandpa is at the table, too, and the grandpa encourages the young man to use the table, so the boy and young man work together, with the boy mostly driving before the young man leaves after a few minutes.*
>
> *Eventually another boy, perhaps 13 years old, arrives and immediately begins manipulating the table simultaneously with the younger boy. Soon, two other children (and their mom) arrive and all four kids (ages maybe 7 – 13 years) are all touching the table at once, first using FloTree then Relate.*[10] *The table crashes several times during this session. The kids talk about what they are choosing for the Relate function.*
>
> *"What's something that doesn't relate? Like this….with this."*
>
> *"Let's see, it could relate!"*
>
> *"Let's hope not. Oh wow!" [The tree returns to the root, so they don't appear to understand that the two species they selected are related, and pick two new species].*

---

[10] Important note: During the first few days of recorded observation, there was a glitch in the table in the Relate function, rendering it erratically unusable. Of the 25 groups that chose the Relate function across the entire observation period, 9 of those groups were unable to get it to work properly. Some persisted for a long while trying to get it to work, with comments like, "*Maybe those [species] aren't related*" when the table wouldn't accept their chosen species. Other visitors did not persist, and some appeared to leave the table in frustration. We account for these software glitches in our analyses below. By the time the Naturalistic study began, this bug had been fixed.

*Then, "It does relate. It relates!" The kids work together like this for some time, although the first boy still dominates much of the activity.*

This boy spends an unusually long time at the exhibit and overlaps with a number of others. However, the set of overlap behaviors is not unusual: Sometimes those arriving at the table while others are there just watch for a while, or watch to try to learn what the exhibit does before engaging. Sometimes they ask questions about how things work and take those who have more experience (even if only a few minutes more) as "experts" (though this may be a pedagogical move on the part of some adults when interacting with children). Sometimes when others arrive, physical or verbal negotiation about what to do ensues, and the activity either becomes more collaborative or sometimes more competitive. Learning what the table does or how it works is an ongoing part of people's interaction with the exhibit, and people are more or less successful at figuring out its features, as described in some of the examples below. Finally, note that despite some of these difficulties, people seem quite interested in exploring with the software – seeking individual species, looking to see if species are related, trying to build separations in FloTree.

*In another case, a young girl of about 5 comes to the table with her dad. Dad reads aloud and encourages her to touch the table. They then pull items from the reel, but do not hold them down long enough to get anywhere except to see the tree zoom a little. They enjoyed the activity, but did not appear to get much information from it. Another couple approached the table, and the girl then "shows" them what to do. This couple ends up also pulling species from the reel but do not hold them long enough to see where it goes. They then use the FloTree but never get into the DeepTree again nor Relate.*

### Characterization of groups

These general behavior patterns vary somewhat with particular group configurations. The following descriptions help us understand some of the *kinds* of behaviors, social interactions, and verbal exchanges that actually occur as visitors engage with the exhibit. It's clear from these observations that visitors are learning both how to interact with the exhibit and about the content displayed, as well as navigating socially with the others at the table. Later, our quantitative summaries of behavior give a broad characterization of what's "typical" in these interactions. These qualitative descriptions provide images of how individuals and groups face particular issues and experience various "Aha!" moments. Since we're trying to capture variation, in this section of the report we try to merge related comments, but sometimes leave somewhat discrepant views side-by-side.

#### *Parents with children*

In one interaction pattern, parents tended to read text aloud to children, both instructions and content, and to focus and direct children to specific areas of the table, pointing and suggesting specific things to do. They more often encouraged children to be drivers, but also did some driving themselves. With some of the younger children (7 or 8 years old) the adult physically guided the child's hand. The children tended to protest this sooner or later. Parents tended to do a lot of talking, asking questions about species, explaining table elements and using expressions like, "*let's try*" and "*what about…*"

Younger kids (ages 4 to 8 or so) appeared to enjoy zooming through the tree with little regard for content.

More than other groups, parent/ child and children only groups were often characterized by using the “find” function (pressing a top image to zoom to it) or pulling species off the reel and holding them to zoom to them, for all or the majority of their time. They would select a species, hold it down until the text became visible, perhaps reading the text or looking at the picture, and then return to select another species, essentially using the tree software in an encyclopedic manner.

Kids often fought over controlling the table, even before they knew what to do. Behaviors included verbally arguing or telling others what to do (or not to do), physically preventing others from touching the table, or simply touching the table to gain control. Because they were mostly pulling organisms from the reel, this meant that they would hold down their organism and move it up or down the table trying to get the tree to move to their organism (i.e., they didn’t understand that the software was zooming just based on holding it down and thought they needed to *do* something to make it work). With multiple kids, there is sometimes lots of simultaneous touching of the table, sometimes making it crash and restart (though it’s not actually clear whether the simultaneous touching caused the crashes we observed). Some kids were more cooperative (sometimes enforced by parents or other adults) and took turns, or negotiated choices (of species from the reel, for example).

*Children alone (not common in the Video Study)*

Kids alone are much more physical with one another than when adults are around. Behaviors include lots of pushing away of hands and holding of arms; jostling for position with bodies; simultaneous touching of the screen, as well as verbal negotiation (“*I want to try this*” and “*Stop*”). Often the older child is more focused on the task and dominant, but there’s some turn-taking.

*Teens*

Teen behaviors can be similar to those of younger kids, but can also be more directed and exploratory. There was one pair of teens, who spent 9:40 at the table. The software crashed three times while they were at the table; it wasn’t clear if that was because of the way they were impatiently touching it or if it was a glitch in the program. They started with FloTree, but they touched the screen so rapidly and briefly they created only a few small barriers, and then the table crashed a couple times within the next minute, so they did not return to FloTree until later during their exploration. They manually explored the DeepTree for a bit and pulled items from the canopy, with initial comments like, “*This is terribly designed. I’ve zoomed it, now what?*” – however, though they seem dissatisfied, it’s not clear what they were expecting the software to do after finding the species. Because both tended to touch the table simultaneously while pulling species from the reel, they found some navigation frustrating and unfruitful, saying, “*We’re just doing nothing.*” Both boys drove at different times and at the same time, and one kept verbally stopping the other, saying, “*Wait wait,*” or “*No*” throughout their whole visit, that is, trying to negotiate cooperative behavior that might have helped them make sense of the possibilities of the exhibit software. Throughout their exploration using Relate, DeepTree, and FloTree, one teen kept alternately saying, “*I get it,*” and then saying, “*No, I don’t.*” Still, despite these difficulties, at one point when zooming using a species from the reel, one boy stops the other from zooming so that he can look more closely at a section of the DeepTree, saying, “*I didn’t realize we could zoom in this far.*” While at first rather unimpressed with the crashing program and their somewhat confusing interactions with the table, they enjoyed and appeared to understand both

FloTree (the second time they used it) and Relate, selecting distantly related species (chameleon and human; praying mantis and ginger).

*Singles*

The sample of twelve "single" adults (5F, 7M) included ten individuals who appeared to be on their own in approaching and leaving the table (but who may have been at the museum with others), as well as two males who had watched or touched the table with a prior group before being at the table alone. The two males who had been part of a larger group spent the least amount of time at the table alone, between 2 and 3 minutes. All the other singles spent from just under 3 to 8-1/2 minutes at the table (median 5:45). One woman approached the table while a child was using the table alone. She watched and left after 3:38 minutes without touching the table. Only one female (at table 5:25) and one male (at table 8:35) explored DeepTree, Relate and FloTree. The other singles only explored one or two of these three main activities.

*Young adults*

Pairs and larger groups of 20-something adults tended to be quite engaged with the table, and conversed about what they were doing and seeing (this was a little harder to generalize because almost half of these used the table when Relate wasn't working right). Dwell time for four pairs ranged from about 13:30 to 24:55 (mean ~17:40). Two of these pairs explored all three main activities (DeepTree, FloTree, and Relate), the other two pairs only moved around in the DeepTree and did not use the FloTree or Relate activities.

A few times it appeared that one or more had studied biology – they referred to it or asked questions of one another. Generally comfortable with content and technology, these groups had lots of conversation about content. There was some simultaneous touching, but unlike teens and children, there was lots of sharing/ turn-taking (one person touching exploring then another seeing something interesting and touching leading the other to pull back.)

One group expressed an interesting novel behavior while exploring FloTree, setting a cooperative group challenge of killing off one set of traits (making it go extinct).

*Older adults*

Some groups of older adults had difficulty using the table, not waiting long enough for a zoom to finish, or all touching the table at once so that it was not responsive.

One group of three was a middle-aged man with an elderly couple. They spent 3:30 at the table. They all touched the table at once, and did not wait to see where it zoomed. Although the woman was able to pull up *Helicobacter Pylori* (the bacteria that causes ulcers) details and pointed out to her husband that this was what lived in his stomach, he responded, "*Big f**ing deal,*" (perhaps he was embarrassed about his condition and didn't want it pointed out). And they left soon after, with the older gentleman commenting, "*It's just a big iPad without the software*."

Another group were three female docents from Cal Academy. They each selected species from the reel, but none waited long enough to reach the end point. They used the FloTree and Relate. They felt they had trouble getting the table to do what they wanted, however, and they remarked that it worked better for younger hands than theirs.

For one couple, the gentleman touches the DeepTree briefly, manually moving up the tree. Thirty seconds into it, the woman says, "*I don't know why you keep doing that,*" and he says, "*I don't know,*" and then touches the Action button. She chooses FloTree. They enjoyed using FloTree and appeared to understand the idea of barriers. But then could not figure out how to get out of FloTree. She asked twice, "*How do we get back?*" and when they couldn't get back, they decided they were done. They spent about 4-1/2 minutes.

A second couple appeared to be slightly older, and they spent almost 29 minutes exploring the table. They tried to use the "Relate" option, but in spite of multiple tries, the table did not accept their choices. They also pulled an organism from the reel, but did not hold it down long enough to get to the end. They moved on to manually explore DeepTree for a long time, moving up and down through the tree. At the very end, 25 minutes into their visit, they chose the Action button again and spent a minute or so using FloTree. The woman remarked a few times that the table made her nauseous (referring to the dizzying effect of flying through the tree).

Some older adults were accompanied by younger adults (perhaps their adult children). In these groups, the younger adult typically drove and also responded to questions from the older adults, and was treated kind of like the "expert" of the group.

In one unusual case, two older adults were accompanied by their teen son. The parents drove the exploration of the table (mostly the man touching the table but the woman pointing out things to try) while their son stood at the side and used his iPad to look up information about particular species that they were encountering on the table – an interesting dynamic.

## Other Notes

There were a handful of times in which a group spent time at the table, and then one group member stayed behind after the others left to have more time to work on the table alone. Sometimes this person was an adult, often one who had been with a child and seemed to want a chance to play themselves; sometimes a child who wasn't quite ready to leave when an adult requested it.

Sometimes with a larger group (three or more), it was difficult for everyone to see or touch what they wanted to, and they would move around the table and switch positions with one another (sometimes in a cooperative manner, sometimes less so).

During FloTree, as noted above, we observed one young adult group create a game. After realizing that colors represented traits, they decided to kill off one set of traits. This pattern of play was also seen in groups of children.

Finally, we present an example of a multi-age grouping including a child, with some typical interaction of adults verbally guiding with the child mostly driving, and of the greater familiarity of younger people with both the technology and the content. They only explore a portion of the table's activities:

> *A family grouping of middle-aged male, middle-aged female and girl about 11 approached the table immediately after another group left, so they did not see any welcome text, just the DeepTree with the origin visible. The girl manually began to move the tree, and then first tapped and then touched, held and expanded an item from the canopy. She continued to slowly manually explore and then the mom said, "Should we read the directions instead of just messing?" She pointed to and read, "Things you can do," and then the girl opened and chose "Experiment." Dad pointed to "What's going on?" and the girl tapped it to open it. Dad then read aloud the text. The girl put her hands down briefly multiple places on the table as the population moved upward.*
>
> *A twenty-something male (who appeared to be part of their party) joined the group at this point and stood by and watched at first. The glowing button for showing the tree shape based on the speciation events appeared, but the girl avoided touching it. Dad said, "Want me to restart it?" and she replied, "No, I want to go back." She tapped the background behind the glowing button, and then the start over button in the lower right corner, but the glowing button just reappeared. She looked at the glowing button again and refrained from touching it.*

*Mom suggested and pointed to the text box at the top of the screen and they all reached for it and opened it. They read it and then Dad suggested, "Let's restart it," pressing the button in the lower right corner. The girl then tapped it multiple times, but it didn't restart, so she tapped the upper right corner box to close, and then, as Dad said, "Do experiment again," the girl tapped buttons to get back into FloTree.*

*Mom said, "Don't go so fast."*

*As they looked at the initial screen, Dad remarked, "So second generation stopped of the blue…somewhere." As the populations started to move upward, the younger man placed his hand on the table and held it there, then the girl did the same, and as the dad also started to put his hand down, he wondered aloud, "Can it take more than one set of hands?"*

*As they looked at their completed FloTree, the girl said, "That looks cool."*

*This time when the option to display the tree form came up, the dad tapped it and the FloTree formed into the tree. They looked at the tree for a moment, and then Dad closed it.*

*Then the younger man pulled a species from the reel, and the girl held it down until the root was visible and the tree paused. She let go and asked, "Want to do it again?" and chose Experiment. Younger man asked, "What about….?" But FloTree was already opening, so they did it again. Dad, young man and girl all put their hands down, and Dad seemed to be trying to read and figure out what was happening.*

*When finished, they looked at the tree form and then back at the population. Dad asked, "So that's…?"*

*Then girl explained, "They are all populations, pink's a population, blue's a population, purple,…they are all separate populations."*

*Then Dad asked, "But one little one turned from pink to blue in this one single strain."*

*Mom asked, "So when there's a barrier, the population stops or something or goes extinct?"*

*At this point, a new little girl approached the table, so this group encouraged her to use the table and they left, with the dad saying, "We failed."*

Despite the final comment, in fact this group seems to have succeeded well in working together to explore the exhibit features, and to learn some of the key concepts in the process. They're curious about how the exhibit works and how to interact with it, and about the meaning of what they're seeing – does it take more than one set of hands to make a barrier? What does it mean when something changes color so quickly? They cooperate in their explorations even though different learners seem to need different speeds and styles of interaction.

### *Quantifying Social Interaction Patterns*

We also conducted a more quantitative examination of observed behavior at the exhibit through the analysis of the "event" data. These "events" refer to coded behaviors that occur within a 20 second time interval experienced collectively by the whole group at the exhibit. We analyzed these data two ways – first focusing on the percent of *time intervals* in which different social events occurred; then focusing on the number of *visitors who experienced the event at least once* while they were at the exhibit. The former approach describes the extent to which certain behaviors were prevalent, overall. However, some behaviors may be rare but important;

thus the latter approach describes the proportion of visitors whose exhibit experiences included them at all.

## DeepTree/ FloTree

The second column in Table 5 shows the proportion of time intervals in which the coded behavior was observed (number of coded intervals divided by total intervals for which person was at the exhibit). The third column in Table 5 shows the proportion of visitors who experienced this behavior during at least one time interval in which they were at the exhibit.

Table 5: Engagement Behaviors for DeepTree Video Study: Proportion over Total Available Intervals and Proportion of visitors who experienced this behavior

| Behavior | Mean proportion of Times Behavior was coded (SD) | Proportion of Visitors who experienced this Behavior |
|---|---|---|
| Pull Person/ Interest | 0.1% (0.8%) | 2% |
| Move Person/ Help | 0.4% (3%) | 4% |
| Prevent Touch/ Control | 5.7% (13%) | 37% |
| Yield | 0.8% (2.7%) | 14% |
| 1 Manipulate | 90% (11%) | 99% |
| Turn Taking | 20% (14%) | 83% |
| 2 Manipulate | 32% (23%) | 88% |
| 3+ Manipulate | 7% (12%) | 42% |
| Point/ Indicate | 23% (16%) | 86% |
| Biology Question | 5% (7%) | 47% |
| Biology Statement | 30% (23%) | 84% |
| Biology Talk Total* | 32% (31%) | 84% |
| How to/ Technical | 23% (18%) | 78% |
| Social Negotiation | 27% (22%) | 80% |
| Read Aloud | 24% (20%) | 82% |
| Refer to Other Time/ Place | 0.5% (2%) | 13% |
| Refer to Other Table Activity | 0.1% (2%) | 15% |
| Unintelligible Talk | 8% (17%) | 35% |
| Enjoy Experience | 7% (11%) | 49% |
| Dislike/ Frustration | 2% (4%) | 24% |
| Relate Glitch | 6% (23%) | 18% |
| Table Crash & Restart | 0.5% (7%) | 8% |

Note: Biology Talk Total includes intervals in which observers coded either a Biology Question, a Biology Statement, or both.
Total intervals: N=1827; Total visitors: N=167.

Notice that sometimes the differences between these ways of looking at social interactions around the exhibit matter a lot. For example, while only 7% of the intervals included someone expressing enjoyment about the exhibit; such an event occurred for nearly half (49%)

of visitors. Similarly, while visitors made connections to other table activities or other experiences in their lives during a minuscule percentage of the time at the table, a full 24% of visitors experienced someone making such a reference during the time they were at the exhibit. Finally, questions or statements about biological content were part of 84% of visitors' experiences, even though they only occurred during about a third of the event intervals.

We also looked at how these behavioral proportions differed by number of people at the table. Tables describing proportion of time intervals are in the Appendices (DeepTree Event Proportions, pp. A-20); those describing proportions of people in each group size who experienced the behavior are in Table 6.

Table 6: Proportion of Visitors who Experienced Engagement Behaviors by Group Size for DeepTree Video Study

| Behavior | 1 person (n = 10) | 2 people (n = 46) | 3 people (n = 52) | 4 people (n = 39) | 5+ people * (n=20) |
|---|---|---|---|---|---|
| Pull Person/ Interest | 0% | 0% | 6% | 0% | 0% |
| Move Person/ Help | 0% | 4% | 6% | 0% | 12% |
| Prevent Touch/ Control | 11% | 24% | 33% | 53% | 59% |
| Yield | 0% | 8% | 29% | 0% | 29% |
| 1 Manipulate | 100% | 96% | 100% | 100% | 100% |
| Turn Taking | 11% | 90% | 82% | 93% | 88% |
| 2 Manipulate | 22% | 86% | 92% | 95% | 100% |
| 3+ Manipulate | 0% | 6% | 45% | 73% | 76% |
| Point/ Indicate | 44% | 88% | 88% | 90% | 88% |
| Biology Question | 0% | 49% | 57% | 50% | 35% |
| Biology Statement | 22% | 76% | 94% | 95% | 82% |
| Biology Talk Total* | 22% | 76% | 94% | 95% | 82% |
| How to/ Technical | 11% | 80% | 92% | 85% | 47% |
| Social Negotiation | 33% | 73% | 86% | 85% | 88% |
| Read Aloud | 11% | 76% | 92% | 93% | 82% |
| Refer to Other Time or Place | 11% | 16% | 18% | 10% | 0% |
| Refer to Other Table Activity | 0% | 18% | 24% | 10% | 0% |
| Unintelligible Talk | 11% | 27% | 41% | 23% | 71% |
| Enjoy Experience | 11% | 63% | 41% | 58% | 29% |
| Dislike/ Frustration | 0% | 22% | 31% | 28% | 0% |

Notes: Biology Talk Total includes intervals in which observers coded either a Biology Question, a Biology Statement, or both. Since only a few people were in groups of 6 or 7, we combine these with groups of 5 to better represent these larger groups.

These group-size related patterns are interesting. While some visitors in groups of 3 or fewer experience "Prevent touch/ control" behaviors, this goes up substantially in groups of 4 or more; a finding which may be related to the extent to which 3 or more people touch the table at once in these larger groups. Also, references to other times or places, or to other table activities;

and expressions of enjoyment as well as dislike or frustration, seem to be more prevalent in groups of 2 to 4 than in larger groups.

We conducted similar observation, coding and analysis of these behaviors for the DeepTree Naturalistic Study (Table 7).

Table 7: Engagement Behaviors for DeepTree Naturalistic Study: Proportion over Total Available Intervals; and Proportion of visitors who experienced this behavior

| Behavior | Mean proportion of times event was coded (SD) | Proportion of Visitors who experienced this event |
|---|---|---|
| Pull Person/ Interest | 2% (10%) | 5% |
| Move Person / Help | 0.9% (4.6%) | 6% |
| Prevent Touch/ Control | 7% (17%) | 22% |
| Yield | 1% (4%) | 10% |
| 1 Manipulate | 92% (20%) | 94% |
| Turn Taking | 10% (19%) | 32% |
| 2 Manipulate | 45% (38%) | 66% |
| 3+ Manipulate | 14% (27%) | 31% |
| Point/ Indicate/ No touch | 13% (19%) | 44% |
| Biology Question | 14% (22%) | 38% |
| Biology Statement | 27% (34%) | 47% |
| Biology Talk Total* | 30% (35%) | 50% |
| How To/ Technical | 33% (34%) | 58% |
| Social Negotiation | 28% (33%) | 57% |
| Read Aloud | 21% (29%) | 45% |
| Refer to Other Time/ Place | 2% (8%) | 13% |
| Refer to Other Table Activity | 0.1% (0.7%) | 2% |
| Unintelligible Talk | 7% (19%) | 17% |
| Enjoy Experience | 15% (24%) | 45% |
| Dislike/ Frustration | 2% (7%) | 12% |

Notes: Biology Talk Total includes intervals in which observers coded either a Biology Question, a Biology Statement, or both.
Total intervals: N=1176; Total visitors: N=326.

Again, large proportions of visitors in the naturalistic study are observed to experience a number of important behaviors — biology talk (50%), enjoyment (45%), and references to other times and places or table activities (15%). Some proportions differ when comparing the video study to the naturalistic study. A generally briefer time at the exhibit in the naturalistic study than the video study gives a smaller chance of observing any particular behavior for each visitor, so the somewhat smaller proportions overall in the naturalistic study are not surprising. Coder differences might also explain differences in some behaviors – e.g., “Turn taking” and “Point/ Indicate” – since Anita and Amy/ Jim differed more in their coding of these behaviors.

In the following tables, we again breakdown behaviors by group size, with the table describing proportion of behaviors by time intervals in the Appendices, DeepTree Event Proportions, pp. A-22, and that showing details by visitors who experienced the behavior in Table 8.

Table 8: Proportion of Visitors who Experienced Engagement Behaviors by Group Size for DeepTree Naturalistic Study

| Behavior | 1 person (n = 48) | 2 people (n = 133) | 3 people (n = 83) | 4 people (n = 24) | 5+ people * (n=28) |
|---|---|---|---|---|---|
| Pull Person/ Interest | 10% | 5% | 6% | 4% | 0% |
| Move Person/ Help | 4% | 4% | 11% | 8% | 0% |
| Prevent Touch/ Control | 4% | 20% | 29% | 21% | 56% |
| Yield | 0% | 8% | 8% | 0% | 60% |
| 1 Manipulate | 94% | 98% | 96% | 100% | 100% |
| Turn Taking | 4% | 38% | 31% | 42% | 60% |
| 2 Manipulate | 19% | 70% | 83% | 83% | 100% |
| 3+ Manipulate | 4% | 14% | 51% | 71% | 84% |
| Point/ Indicate | 15% | 53% | 42% | 58% | 68% |
| Biology Question | 10% | 50% | 29% | 33% | 84% |
| Biology Statement | 15% | 62% | 39% | 42% | 84% |
| Biology Talk Total* | 17% | 65% | 45% | 50% | 84% |
| How to/ Technical | 25% | 68% | 60% | 63% | 84% |
| Social Negotiation | 23% | 65% | 64% | 67% | 84% |
| Read Aloud | 21% | 53% | 42% | 42% | 88% |
| Refer to Other Time or Place | 2% | 18% | 11% | 4% | 24% |
| Refer to Other Table Activity | 0% | 2% | 1% | 13% | 0% |
| Unintelligible Talk | 8% | 21% | 18% | 25% | 16% |
| Enjoy Experience | 21% | 50% | 45% | 63% | 72% |
| Dislike/ Frustration | 2% | 12% | 11% | 46% | 12% |

Notes: Biology Talk Total includes intervals in which observers coded either a Biology Question, a Biology Statement, or both. Since only a few people were in groups of 6 or 7, we combine these with groups of 5 to better represent these larger groups.

## BAT studies

We conducted similar analyses for behaviors around the Build-A-Tree exhibit, again looking both at the proportion of time intervals in which different behaviors occurred, and the proportion of visitors who experienced each behavior during some portion of their time at the exhibit. Remember that the BAT studies were smaller than the DeepTree studies, and the behavioral events coded were designed with the DeepTree study in mind, though we felt they were sufficiently typical of what might be seen around the BAT exhibit to use them there as well. As before, we present overall statistics and a breakdown by group size of proportion of visitors who experienced each behavior in Table 9 and Table 10, and a breakdown by group size of proportion of intervals in which an event occurred in the Appendices (BAT Event Proportions, pp. A-22 to A-24).

Table 9: Proportion of Engagement Behaviors over Total Available Intervals BAT Video Study

| Behavior | Mean proportion of times event was coded (SD) | Proportion of Visitors who experienced this event |
|---|---|---|
| Pull Person/ Interest | | |
| Move Person / Help | | |
| Prevent Touch/ Control | | |
| Yield | | |
| 1 Manipulate | 93% (10%) | 100% |
| Turn Taking | 17% (15%) | 78% |
| 2 Manipulate | 22% (20%) | 78% |
| 3+ Manipulate | 3% (6%) | 17% |
| Point/ Indicate | 26% (23%) | 70% |
| Biology Question | 10% (7%) | 78% |
| Biology Statement | 36% (28%) | 83% |
| Biology Talk | 40% (28%) | 83% |
| How To/ Technical | 14% (12%) | 83% |
| Social Negotiation | 22% (17%) | 78% |
| Read Aloud | 28% (20%) | 83% |
| Refer to Other Time/ Place | 1% (1%) | 17% |
| Refer to Other Table Activity | 0.2% (0.6%) | 17% |
| Unintelligible Talk | 5% (5%) | 61% |
| Enjoy Experience | 8% (6%) | 70% |
| Dislike/Frustration | 0.4% (0.9%) | 22% |

Note: Biology Talk Total includes intervals in which observers coded either a Biology Question, a Biology Statement, or both.
Total intervals: N=407; Total visitors: N=23.

In the BAT Video Study, as in the DeepTree Video Study, large and comparable proportions of visitors experience someone manipulating the table, engaging in biology-related talk, reading text from the exhibit, and making references to other events in their lives. Slightly higher percentages experience someone expressing enjoyment of their experience in the BAT Video Study (70%) than the DeepTree Video Study (49%) though it's not clear if that difference is significant.

Table 10: Proportion of Visitors who Experienced Engagement Behaviors by Group Size for BAT Video Study

| Behavior | 1 person (n = 5) | 2 people (n = 12) | 3 people (n = 2) | 4 people (n = 4) |
|---|---|---|---|---|
| Pull Person/ Interest | | | | |
| Move Person / Help | | | | |
| Prevent Touch/ Control | | | | |
| Yield | | | | |
| 1 Manipulate | 100% | 100% | 100% | 100% |
| Turn Taking | 20% | 100% | 50% | 100% |
| 2 Manipulate | 20% | 100% | 50% | 100% |
| 3+ Manipulate | | | | 100% |
| Point/ Indicate/ No touch | | 100% | | 100% |
| Biology Question | 20% | 100% | 50% | 100% |
| Biology Statement | 40% | 100% | 50% | 100% |
| Biology Talk | 40% | 100% | 50% | 100% |
| How To/ Technical | 40% | 100% | 50% | 100% |
| Social Negotiation | 20% | 100% | 50% | 100% |
| Read Aloud | 40% | 100% | 50% | 100% |
| Refer to Other Time/ Place | | 33% | | |
| Refer to Other Table Activity | | 33% | | |
| Unintelligible Talk | | 83% | | 100% |
| Enjoy Experience | 40% | 83% | | 100% |
| Dislike/Frustration | 20% | 33% | | |

Note: Biology Talk Total includes intervals in which observers coded either a Biology Question, a Biology Statement, or both.

As before, we conduct similar analyses for the BAT Naturalistic Study, first presenting overall proportions by time intervals and proportion of visitors experiencing a behavior (Table 11), then breaking these down by group size (Table 12), with the data for proportion of intervals in which an event occurred in the Appendices (BAT Event Proportions, p. A-22).

Table 11: Engagement Behaviors for BAT Naturalistic Study: Proportion over Total Available Intervals; and Proportion of visitors who experienced this behavior

| Behavior | Mean proportion of times event was coded (SD) | Proportion of Visitors who experienced this event |
|---|---|---|
| Pull Person/ Interest | 2% (9%) | 11% |
| Move Person / Help | 4% (10%) | 25% |
| Prevent Touch/ Control | 8% (15%) | 34% |
| Yield | 0.7% (3%) | 8% |
| 1 Manipulate | 94% (19%) | 97% |
| Turn Taking | 8% (12%) | 43% |
| 2 Manipulate | 45% (35%) | 75% |
| 3+ Manipulate | 8% (21%) | 25% |
| Point/ Indicate | 14% (18%) | 50% |
| Biology Question | 15% (21%) | 48% |
| Biology Statement | 32% (33%) | 63% |
| Biology Talk* | 34% (34%) | 65% |
| How To/ Technical | 36% (31%) | 72% |
| Social Negotiation | 20% (23%) | 65% |
| Read Aloud | 27% (30%) | 58% |
| Refer to Other Time/ Place | 1% (4%) | 15% |
| Refer to Other Table Activity | | 0% |
| Unintelligible Talk | 7% (20%) | 17% |
| Enjoy Experience | 8% (13%) | 43% |
| Dislike/Frustration | 4% (12%) | 20% |

Note: Biology Talk Total includes intervals in which observers coded either a Biology Question, a Biology Statement, or both.
Total intervals: N=689; Total visitors: N=155.

Naturalistic visitors at the BAT exhibit experience expressions of enjoyment, biology talk, and references to other times and places that are comparable to both BAT Video Study visitors, and DeepTree Naturalistic visitors. Somewhat larger proportions experience "Move person/ Help" behaviors than in any of the other studies or conditions.

Table 12: Proportion of Visitors who Experienced Engagement Behaviors by Group Size for BAT Naturalistic Study

| Behavior | 1 person (n = 24) | 2 people (n = 50) | 3 people (n = 47) | 4 people (n = 18) | 5 people (n = 6) | 6 people (n = 10) |
|---|---|---|---|---|---|---|
| Pull Person/ Interest | 8% | 16% | 13% | 6% | 0% | 0% |
| Move Person / Help | 8% | 24% | 19% | 17% | 50% | 100% |
| Prevent Touch/ Control | 17% | 40% | 36% | 11% | 83% | 40% |
| Yield | 4% | 12% | 6% | 0% | 0% | 30% |
| 1 Manipulate | 96% | 100% | 100% | 83% | 100% | 100% |
| Turn Taking | 17% | 50% | 38% | 44% | 83% | 60% |
| 2 Manipulate | 21% | 80% | 87% | 83% | 83% | 100% |
| 3 Manipulate | 4% | 6% | 45% | 11% | 83% | 60% |
| Point/ Indicate | 21% | 52% | 51% | 44% | 83% | 100% |
| Biology Question | 25% | 56% | 45% | 33% | 83% | 90% |
| Biology Statement | 33% | 72% | 57% | 56% | 100% | 100% |
| Biology Talk | 33% | 72% | 64% | 61% | 100% | 100% |
| How To/ Technical | 29% | 82% | 77% | 67% | 83% | 100% |
| Social Negotiation | 33% | 72% | 70% | 44% | 83% | 100% |
| Read Aloud | 25% | 64% | 53% | 61% | 100% | 100% |
| Refer to Other Time/ Place | 4% | 18% | 15% | 17% | 50% | 0% |
| Refer to Other Table Activity | 0% | 0% | 0% | 0% | 0% | 0% |
| Unintelligible Talk | 4% | 20% | 19% | 0% | 0% | 60% |
| Enjoy Experience | 17% | 44% | 53% | 22% | 83% | 60% |
| Dislike/ Frustration | 13% | 32% | 17% | 17% | 17% | 0% |

Note: Biology Talk Total includes intervals in which observers coded either a Biology Question, a Biology Statement, or both.

### *Table Activity Data*

While the observation data provides a description of social interaction around the table, the table activity data provide specific measures of how the DeepTree/FloTree exhibit is being used, and which elements of the software are being manipulated, for how long and/or how many times. (These analyses only apply to the DeepTree video study, as log data were not collected for the Build-a-Tree exhibit, and we were not able to synchronize table log measures with our other data in the DeepTree naturalistic study.) As noted above, the evaluation team worked with the development team during the spring, 2013, to define 78 behavioral measures of table activity that could be derived from the low level logs. These included summaries for each person of time spent doing different activities, number of times that certain behaviors or actions occurred, or summary statistics (average, standard deviation) of relevant "distances" describing extent of coverage of the Tree of Life, or degree of relatedness of several species, or distance on the species reel. Measures were defined to characterize all the major activities available at the exhibit:

- Manual navigation.
- Top image navigation, in which a visitor holds an image at the canopy and the software flies towards it.
- Reel item navigation, in which a visitor pulls a species off the species reel and holds it, and the software flies towards that species.
- Inspection of text and top image zooming, describing ways in which visitors pay closer attention to information about various species.
- Trait display, available via a flashing icon at some important nodes. Pressing the icon opens a description of the key trait represented by this branching point in the Tree of Life, with some pictures and text for examples of species that represent it.
- Relate, in which visitors pull pairs of species off the image reel, and the software flies towards the point showing the two species and their most recent common ancestor. As noted previously, there were problems with the functioning of Relate for 18% of visitors in the DeepTree Video study, but this bug was fixed before the Naturalistic study.
- Training tree, a simplified tree diagram available after completing a Relate search, showing major speciation points and shared lineage of the related species.
- FloTree, or Experiment, an interactive mode depicting the process of speciation; in particular, how populations with inherent variation, when separated by physical barriers, can lead to new species.

A full description of all these measures, along with summary statistics (mean, standard deviation, median, minimum and maximum) of their occurrence appears in the Appendices Table Log Summary Measures and Summary Statistics for Table Log Measures, pp. A-24 to A-36.

These descriptives suggest that people were engaged with many of the available functions in the software. Serrell (1997) suggests a measure of exhibit engagement is the "percentage of diligent visitors (%DV)" — those who stop at/ attend to at least half of an exhibition's exhibits. For DeepTree, we try to capture this by calculating the number of activities/ features that each visitor attended to, even if only minimally. (Serrell also suggests a "Sweep Rate Index (SRI)" which is the exhibition square footage divided by total time spent, with lower numbers suggesting a longer dwell time – but this statistic doesn't have a clear analogue with DeepTree.)

Table 13: Proportion of DeepTree Video Study Visitors Engaging in Different Numbers of Activities

| Number of Activities (beyond Manual Navigation) | Percentage of Visitors Engaging in This Many Activities | Percentage of Visitors Engaging in *at Least* This Many Activities |
|---|---|---|
| 1 | 4.2% | 100.0% |
| 2 | 12.6% | 95.8% |
| 3 | 24.0% | 83.2% |
| 4 | 21.6% | 59.3% |
| 5 | 14.4% | 37.7% |
| 6 | 11.4% | 23.4% |
| 7 | 12.0% | 12.0% |

Excluding manual navigation, which occurred as soon as a visitor touched the table, there are seven (7) primary activities. Table 13 displays percentages of visitors engaging in different numbers of activities. Note that 38% of DeepTree Video study participants engaged in 5 or more of the 7 activities, and 59% of visitors engaged in 4 or more activities – a high proportion of "Diligent Visitors."

Table 14 provides summary statistics on selected measures. A few things to note:

1) It's not clear how to think about the measure of unique focal nodes visited (NAV COUNT UNIQUE FOCAL NODES). The median value is 117, which seems like a fair number of different species to view while navigating the tree. While there were about 70,000 species visible in the tree, the "focal node" is the largest visible internal node, and only about 20,000 species are possible "focal nodes." The median of 117 is about 0.6 of

Table 14: Selected Table Measure Descriptives

| Measure | Mean (SD) | Median | Min-Max |
|---|---|---|---|
| Amount of time spent using any of the navigation types (seconds). NAV_TIME_NAVIGATING | 116.79 (176.39) | 57.19 | 0-1168 |
| Total number of unique focal nodes visited. NAV_COUNT_UNIQUE_FOCAL_NODES | 118.09 (79.82) | 117 | 0-368 |
| Time spent manually navigating (seconds). NAV_M_TIME_SPENT | 81.75 (107.64) | 53 | 0-668 |
| Number of species pulled out from the reel during free exploration. NAV_R_COUNT_SPECIES_PULLED_OUT | 5.36 (6.92) | 3 | 0-31 |
| Number of seconds spent holding an image reel item (seconds). NAV_R_TIME_SPENT | 89.85 (124.86) | 19.37 | 0-536 |
| Number of unique reel items held at least once. NAV_R_COUNT_SPECIES_HELD | 5.09 (6.07) | 3 | 0-25 |
| Of the n= 119 who held any reel items: | 7.24 (6.08) | 6 | 1-25 |
| Number of unique species that became visible while corresponding reel item was held NAV_R_COUNT_SPECIES_NAVIGATED_TO | 2.33 (3.87) | 0 | 0-15 |
| Of the n= 119 who held any reel items: | 3.29 (4.26) | 2 | 0-15 |
| Number of seconds spent holding a top image (seconds). NAV_T_TIME_SPENT | 15.00 (20.10) | 5.93 | 0-133 |
| Number of unique images held at least once. NAV_T_COUNT_SPECIES_HELD: | 4.82 (6.95) | 3 | 0-48 |
| Of the n= 133 who held any top images: | 6.13 (7.30) | 4 | 1-48 |
| Number of unique species that became visible while corresponding top image was held. NAV_T_COUNT_SPECIES_NAVIGATED_TO | 1.06 (1.45) | 0 | 0-6 |
| Of the n= 133 who held any top images: | 1.34 (1.52) | 1 | 0-6 |
| Number of species text that are visible on screen for at least 10 seconds. INS_COUNT_TEXT_10_SEC | 4.8 (5.07) | 3 | 0 -27 |
| Time spent in the Trait Display (seconds). TRAIT_TIME_SPENT | 60.23 (103.52) | 23.97 | 0-738 |
| Number of times that the Trait display is activated TRAIT_COUNT | 2.45 (2.92) | 2 | 0-20 |

Note: N=169

1% of all possible focal nodes – a small fraction of the total. Of course, we never expect people to view all the species. What is a reasonable expectation for a brief exhibit visit? Did visitors meet it?

2) While people seemed to try Top Level Navigation and Reel Item Navigation, the majority didn't succeed in navigating all the way to the species being held (medians for both = 0). However, limiting the statistic just to those who attempted each type of navigation at least once, we find that median number of attempts is slightly higher, and also that people succeed at least a little bit.

Dichotomous variables were created to represent whether people "found" the Relate function and FloTree. About 32% of subjects found the Relate function (n = 54) and 52% of participants found FloTree (n = 89). There was a significant correlation between those who found Relate and those who found FloTree ($r = .52$, $p < .001$) likely because these were the primary options on the Action button. Therefore, it was often the case that, if individuals found Relate, they also spent some time in FloTree, or vice versa. But, this was not always the case.

Table 15 lists a subset of descriptive statistics for the Relate function for those people who found it (full descriptive statistics appear in the Appendices, Summary Statistics for Table Log Measures, p. A-30 to A-34). Because Training Tree was only available once subjects found the Relate function, we've also pulled out descriptive statistics on Training Tree activity for those individuals who found this function.

Table 15: Selected Relate Function Descriptives

| Measure | Mean (SD) | Median | Min-Max |
|---|---|---|---|
| Number of relate queries executed. REL_COUNT | 2.91 (2.88) | 2 | 1–17 |
| Number of species pulled out from the reel while selecting species. REL_COUNT_SPECIES_PULLED_OUT | 5.48 (6.55) | 4 | 0–39 |
| Time spent selecting species while the dialog is active (seconds). REL_TIME_SPENT_SELECTING_SPECIES | 71.16 (59.62) | 67.89 | 9–386 |
| Average time spent per trait across all Trait Displays (seconds). TT_ AVRG_TIME_SPENT_PER_TRAIT | 8.22 (8.66) | 7.30 | 0–34 |
| Number of traits selected in the Training Tree across all Trait Displays. TT_TRAIT_COUNT | 10.04 (15.31) | 5.50 | 0-80 |

Note: Subsample (n = 54) of those who found Relate.

People who found the Relate function seemed to spend a fair amount of time selecting species to work with it (though some of this time may have been spent trying to get Relate to work when the software was experiencing a bug.)

We also created a set of descriptive statistics for the FloTree function for those people who found it (Table 16).

Again, it seems that those who found the FloTree function were successful in running it, and creating a number of species. They also seemed to use the function to transform the population diagram to the tree diagram, which was intended to provide scaffolding for linking the FloTree representation to the DeepTree representation, and which the design team incorporated as a more automatic feature after evaluation testing was complete.

Table 16: Selected FloTree Descriptives

| | Mean (SD) | Median | Min-Max |
|---|---|---|---|
| Number of times FloTree/ Experiment was activated. FT_COUNT | 1.42 (0.75) | 1 | 1-5 |
| Time spent in the FloTree (seconds). FT_TIME | 169.61 (97.04) | 155.47 | 11 – 586 |
| Total number of experiments launched. FT_COUNT_EXPERIMENTS | 1.87 (1.18) | 2 | 0-8 |
| Number of experiments with at least one speciation event. FT_COUNT_SPECIATION_EXPERIMENTS | 1.56 (1.11) | 1 | 0-6 |
| Total number of species created in experiments. FT_COUNT_SPECIES | 8.51 (7.28) | 7 | 0-40 |
| Average number of species created per experiment. FT_AVRG_SPECIES | 4.19 (2.26) | 4 | 0-9.33 |
| Number of times they transform population view to tree view. FT_COUNT_TREE_DIAGRAM | 10.01 (8.82) | 7 | 0-56 |

Note: Subsample (n = 89) of those who found FloTree/ Experiment.

### *Qualitative Examination of Table Log Files*

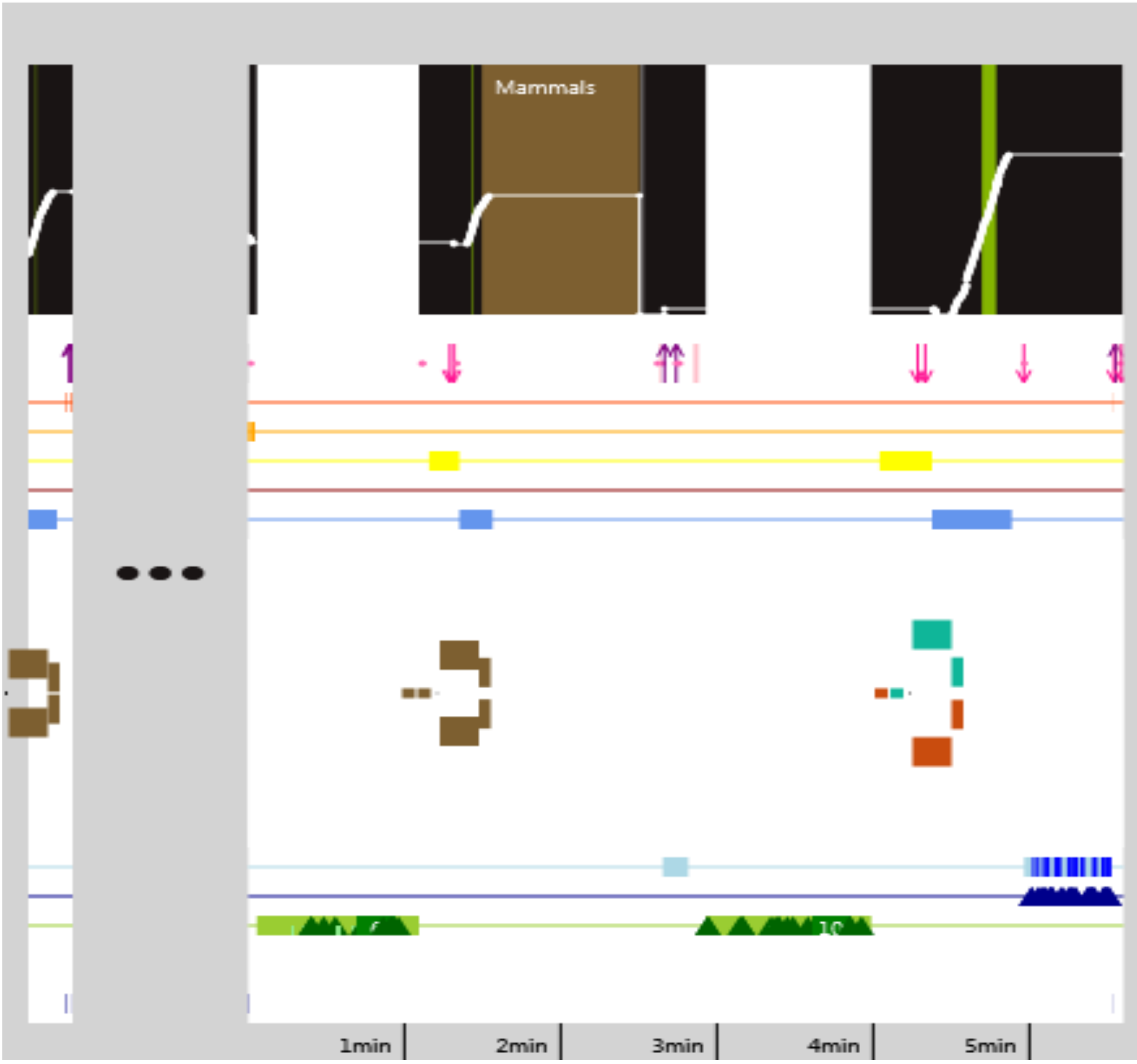
The development team used the table log files to create graphical representations of each visitors' experience with DeepTree. Individual user log files were examined by the evaluation team in order to gain more understanding of users' experience with the table (see Appendix, Semantic Log Visualization Guide, p. A-37). Below we examine and discuss some of the ways in which visitors used the exhibit based on the log files that were collected during their visit. The experience of User 297 who demonstrated a high level of understanding and used appropriate terms to describe the exhibit in the open-ended survey question, may be illustrative of a full experience at the exhibit.

In response to the survey question asking what the exhibit was about, this user explained the exhibit as being an "*[o]verview of evolution and demonstration of how seemingly very diverse animals might actually share a common ancestor.*" This response showed that the user grasped the key concepts of the exhibit, but the user did not already have a degree in biology as many of the others who demonstrated high levels of understanding did, and thus her experience can highlight the experience of someone who gained from the exhibit.

User 297 experienced the exhibit for approximately six minutes, somewhat less than the median for the DeepTree Video Study. She started to work with relate and did some manual navigation, then left the table (the grey gap in the Log Visualization, Figure 18). Once she returned, she opened the FloTree option (the white gap at top and green icons below) and created multiple speciation events. After having completed the FloTree, she moved on to the Relate function where she compared two mammals, then looked at the Trait display. She later explored manually through the tree, used the FloTree a second time as well as the Relate function where she compared a bird and a dinosaur. Toward the end of her experience she examined Traits in the training tree. This experience, although not exploring much breadth of the tree, does show a deep engagement with the exhibit. The user experiences many of the teaching features such as FloTree, Relate and Trait display. As a result it is not surprising that she would demonstrate a high level of understanding of the exhibit.

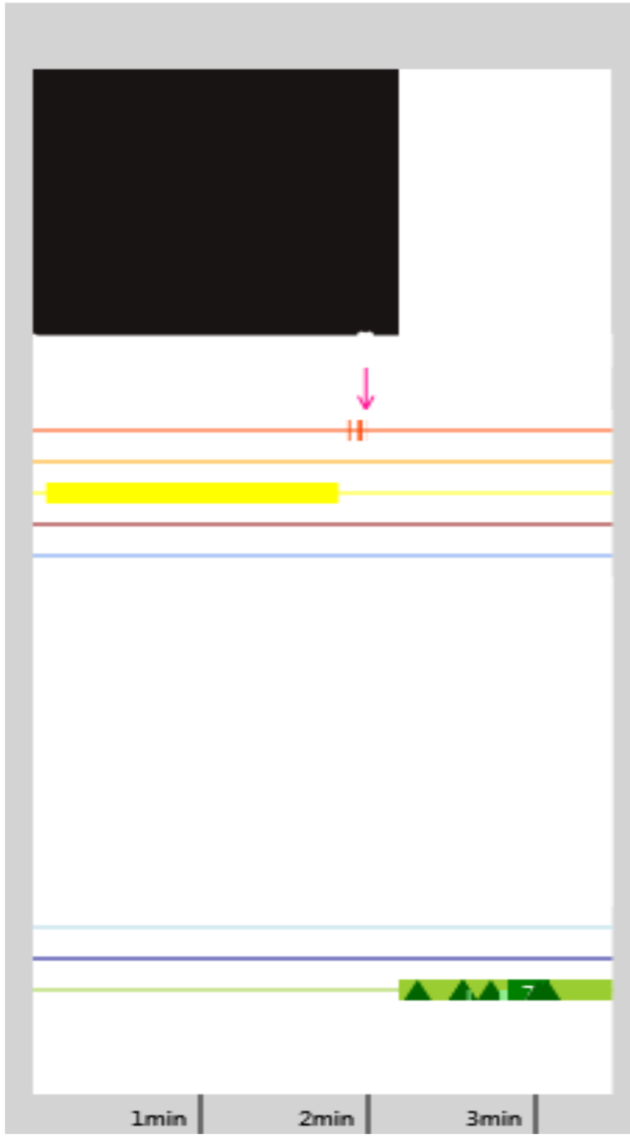
On the other hand, User 238 demonstrates a lower level of understanding of the exhibit, stating "*I'm not sure,*" when asked what it's about and, interestingly, his experience at the table is very limited. As can be seen in the chart (Figure 19), the user experienced far less of the exhibit and for less time, spending approximately four minutes at the exhibit. There is very little

Figure 18: Log file visualization: User 297




depth or breadth in the experience of this user who seemed to mostly hold a top image in the canopy of the tree and hold a species from the reel and then move on to the FloTree. There was almost no manual navigation, the variety of species viewed is limited, and the Relate function is never used. During use of the FloTree the user is able to create speciation events, but then ends his experience shortly after. It is not clear whether a lack of interest or engagement resulted in less learning, or whether pre-existing knowledge or interest would lead to deeper engagement with the table. Nevertheless, these two examples highlight two different types of use of the exhibit. User 297, although she did not spend a large amount of time at the exhibit, did spend enough time to interact with all of the elements and to explore some of the tree manually. On the other hand User 238 was less engaged and explored the tree very little and failed to activate the Relate function, which is most explicit in explaining the key concept of common ancestry.

Figure 19: Log file visualization: User 238




Although these two users' post-exhibit responses illustrate a high level of understanding on the one hand and a low level on the other, we cannot say for sure that their experience at the table was solely responsible for their demonstrated levels of understanding. Instead they do exemplify two distinct types of use and levels of understanding.

### *Exhibit Engagement Factor Analysis*

While we have been able to describe the myriad ways that visitors engaged with the touch table exhibit and with one another, we hoped to consolidate the information from the large number of variables collected into a single composite engagement score. The intention was to capture time spent, and extent of exploration with the various software tools provided to be able to say, no matter *how* a specific visitor interacted with the exhibit, that they were more or less engaged with its content. While we already had information about the amount of time people spent at the table, we believed that, by adding other measures to the calculation, a richer assessment of engagement could be obtained. We attempted to use confirmatory factor analysis

(CFA) to develop a latent factor score that would represent such an engagement variable. However, our models had poor overall fit and low reliability in several of the indicator measures such that we didn't feel confident using these in our broader analyses. An encouraging aspect of the factor analysis was that general time spent was the most reliable and valid indicator of engagement. Therefore, we believe that it is reasonable to use time spent at the exhibit as a proxy for engagement, since a more complex measure was found to be premature. Details of our factor analytic exploration can be found in the Appendix, Confirmatory Factor Analysis of Exhibit Engagement, pp. A-34 to A-37.

## *Survey Responses*

### Open-Ended Survey Results

The open-ended responses from the question "What was the touch table about?" were grouped by keywords/ themes that emerged from the answers (Table 17). The most common phrase used to describe the exhibit when it was running the DeepTree software was "*Tree of Life*." This included writing the phrase "*Tree of Life*" or something similar or explaining the tree of life in more detail. This is not surprising considering the name of the exhibit was Tree of Life. Other common descriptions used when running the DeepTree software were "*evolution*" and "*relatedness*." Although none of these indicate that respondents had a deep understanding of the concepts, it does indicate a degree of basic understanding of the exhibit.

On the other hand when the exhibit table was running the BAT software respondents were far less likely to describe the exhibit as being about the Tree of Life or evolution. Instead they were more likely to describe the exhibit as being about "*relatedness*." This was the only word (or related word) that was used consistently to describe the exhibit regardless of which software was running. It seems based on these responses then, that users had a much different conception of what the exhibit was about depending on the software they were using.

Less common ways of describing the exhibit table when it was running the DeepTree software, but which also exhibited understanding, included describing the exhibit as being about "*Time*," "*Common Ancestry*," "*Connectedness*," "*DNA*," "*Origins of Life*," "*Barriers/obstacles*" and "*Origins of Life*."

For those who used the BAT software, aside from describing the exhibit as being about relatedness, the next most common ways to describe the exhibit included "*connectedness*," "*phylogenetics*," and "*grouping*." It should be noted that most of those who described the exhibit as being about phylogenetics had degrees in biology or an advanced degree. Several children described the exhibit as being about "*grouping*."

After grouping the responses by keyword/ theme, we examined them for extent of understanding of the exhibit content. The responses varied greatly in the degree of understanding displayed. While a portion of the responses were relatively complete and used scientific terms, for example, "*Origin of species/common ancestry. History of life, relationship of species*," most responses were more limited in scope, reflecting just one aspect of the exhibit ("*It was about how genetics and barriers create new species and wipe out others*") or providing a much briefer explanation ("*Evolution*," "*How animals are connected*"). Most frequently, users demonstrated that they identified the key concepts of the exhibit but did not use precise or scientific terms to describe it. This appeared to be true for both the BAT and the DeepTree exhibits, and across video and naturalistic studies. A much smaller portion of visitors' responses demonstrated a fundamental lack of understanding of the exhibit ("*learning history in a fun way*," "*not sure*").

Table 17: List of Keywords/ Themes for LOE Survey responses

| Keyword/ Theme | DeepTree Video (n=132) | DeepTree Naturalistic (n=33) | DeepTree Totals (n=165) | BAT Video (n=19) | BAT Naturalistic (n=12) | BAT totals (n=31) |
|---|---|---|---|---|---|---|
| Tree of Life | 39 (30%) | 15 (47%) | 54 (33%) | 1 (5%) | 1 (7%) | 2 (6%) |
| Evolution | 32 (24%) | 11 (33%) | 43 (26%) | 2 (11%) | 0 (0%) | 2 (6%) |
| Relatedness (includes Relations, Relationship, Relatives, Relativity, Common Relations) | 35 (27%) | 6 (18%) | 41 (25%) | 9 (47%) | 7(58%) | 16 (51%) |
| Time | 11 (8%) | 0 (0%) | 11 (7%) | 0 (0%) | 0 (0%) | 0 (0%) |
| Common Ancestry | 6 (5%) | 2 (6%) | 8 (5%) | 0 (0%) | 0 (0%) | 0 (0%) |
| Connectedness (includes Connecting, Connections, Connected) | 4 (3%) | 3 ( 9%) | 7 (4%) | 2 (11%) | 0 (0%) | 2 (6%) |
| DNA | 6 (5%) | 2 (9%) | 8 (5%) | 0 (0%) | 0 (0%) | 0 (0%) |
| Origins (of Life) | 8 (6%) | 1 (3%) | 9 (5%) | 0 (0%) | 0 (0%) | 0 (0%) |
| Barriers/Obstacles | 5 (4%) | 2 (9%) | 7 (4%) | 0 (0%) | 0 (0%) | 0 (0%) |
| Phylogenic/ phylogenetic | 2 (2%) | 0 (0%) | 2 (1%) | 4 (21%) | 1 (8%) | 5 (16%) |
| Fun | 2 (2%) | 1 (3%) | 3 (2%) | 0 (0%) | 0 (0%) | 0 (0%) |
| Family Tree | 0 (0%) | 2 (9%) | 2 (1%) | 0 (0%) | 1 (8%) | 1 (3%) |
| Grouping | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 2 (17%) | 2 (6%) |
| History of Life | 3 (2%) | 0 (0%) | 3 (2%) | 0 (0%) | 0 (0%) | 0 (0%) |
| Adaptation | 1 (1%) | 1 (3%) | 2 (1%) | 0 (0%) | 0 (0%) | 0 (0%) |

## Learning Outcomes

In addition to demographic information, ratings of interest and enjoyment, and the open-ended responses reported above, visitors completing the survey responded to several questions from the Learning Research of the *Life on Earth* project. Using 5-point Likert scales, they rated their level of agreement with questions about relatedness of various species (1=disagree a lot, 5 = agree a lot), and with several statements about evolution (1=disagree a lot, 5 =agree a lot).

The questions about common ancestors all had the same structure, asking about level of agreement with the statement: "SPECIES A and SPECIES B had the same ancestor a long, long time ago." Table 18 reports results for the four survey questions and the average of all of them (see Appendix, DeepTree Survey Results, p. A-41 for further details).

Respondents tended to agree with all these statements, though they more strongly agreed that mice and rats had the same ancestor, and somewhat more that rabbits and lizards did than the more disparate species.

Visitors also rated their level of agreement with the statements about evolution (Table 19), as well as the average of all of them. Here they had very high levels of agreement with all the statements (see Appendix, DeepTree Survey Results, p. A-41 for further details).

Table 18: Levels of Agreement with Survey Questions about Common Ancestry of Various Species.

| Species Pairs | DeepTree Video | DeepTree Naturalistic | BAT Video | BAT Naturalistic |
|---|---|---|---|---|
| Rabbits and Lizards | 3.88 (1.16) | 3.77 (1.45) | 3.89 (1.24) | 3.09 (1.22) |
| Humans and Mushrooms | 3.23 (1.53) | 3.63 (1.61) | 3.11 (1.78) | 2.42 (1.44) |
| Mice and Rats | 4.51 (0.85) | 4.43 (1.10) | 4.63 (0.60) | 4.50 (0.90) |
| Bears and Sunflowers | 3.27 (1.54) | 3.37 (1.73) | 2.89 (1.66) | 2.42 (1.44) |
| **Common Ancestor Composite** | **3.71 (1.05)** | **3.80 (1.21)** | **3.65 (1.05)** | **3.09 (0.91)** |

Note: All cells show Mean (SD).
DeepTree Video N=125. DeepTree Naturalistic N=30.
BAT Video N=19. BAT Naturalistic N=12.

When looking at the composite scores, on average, across all ages and educational backgrounds, participants rated the common ancestry questions 3.71 (*SD* = 1.05) and rated the evolution questions 4.23 (*SD* = .77). While it is not possible to know what knowledge people brought to the exhibit prior to the experience, this indicates that subjects, when surveyed after their experience, were more likely to agree with these statements about evolution than the statements about common ancestry. The correlation between individuals' common ancestry ratings and their evolution ratings was $r = .43$ ($p < .01$).

Table 19: Levels of Agreement with Survey Questions about Evolution

| Statement | DeepTree Video | DeepTree Naturalistic | BAT Video | BAT Naturalistic |
|---|---|---|---|---|
| CARDINALS, a type of bird, are changing over time. They might be VERY different millions of years in the future. | 4.23 (0.93) | 4.16 (1.00) | 4.68 (0.48) | 3.33 (1.30) |
| Most living things today are VERY different from their ancestors who lived a long, long time ago. | 4.07 (1.03) | 4.19 (1.22) | 4.32 (1.00) | 3.75 (1.60) |
| HUMAN BEINGS, a type of primate, are changing over time. They might be VERY different millions of years in the future. | 4.01 (1.07) | 4.29 (0.86) | 4.42 (0.96) | 3.83 (1.40) |
| Evolution is still going on TODAY. | 4.66 (0.73) | 4.61 (0.67) | 4.63 (0.96) | 4.42 (1.16) |
| COYOTES, a type of mammal, are changing over time. They might be VERY different millions of years in the future. | 4.19 (0.97) | 4.13 (0.99) | 4.63 (0.50) | 3.83 (1.27) |
| **Evolution Composite** | **4.23 (0.77)** | **4.28 (0.78)** | **4.54 (0.51)** | **3.83 (1.18)** |

Note: All cells show Mean (SD).
DeepTree Video N=125. DeepTree Naturalistic N = 31.
BAT Video N=19. BAT Naturalistic N = 12.

We conducted similar analyses for the DeepTree Naturalistic Study participants. Across the 30 DeepTree/FloTree survey participants in the Naturalistic Study, participants rated common ancestry composite items 3.80 (*SD* = 1.21) overall, and evolution composite items 4.28 (*SD* = .78). This difference again suggests that the small number of survey respondents in the Naturalistic study also agreed more with the statements about evolution than with the statements about common ancestry. The correlation between common ancestry and evolution ratings is about the same, $r = .42$ ($p = .02$).

For the BAT study, we conducted similar analyses (see Table 18 and Table 19 and Appendix, BAT Survey Results, p. A-43 for further details). Again, we find a similar pattern of results, though visitors somewhat disagreed that bears and sunflowers had a common ancestor. Overall, levels of agreement with the evolution composite are again significantly higher than agreement with the common ancestry composite. Correlation between these composites is lower and no longer statistically significant ($r = .21$, $p = .40$) but the sample size is small which could affect significance levels.

With the BAT Naturalistic Study, we find similar results (see Table 18 and Table 19 and Appendix, BAT Survey Results, p. A-43 for further details): Agreement with all the common ancestry questions (and stronger agreement with mice and rats); and higher levels of agreement on the evolution questions, with a statistically significant correlation between them, $r = .48$, $p = .001$.

## *Relationships Among Variables*

We now turn to an examination of relationships among the several types of data collected. We're especially interested in ways that demographic characteristics, group size, or experience at the exhibit table are related to engagement or learning outcomes. A few caveats: While we are looking for relationships among variables, these are only observational data and therefore we can't say that any relationships found are causal. Also, though we attempted to find a more inclusive "Engagement score" through confirmatory factor analysis, we weren't able to create a reliable index and so returned to the proxy variable of dwell time. Finally, our ability to explore relationships among variables is limited by the extent to which we were able to link variables – therefore we explore more relationships with the DeepTree video data; fewer with the other studies and data sets.

We begin with an exploration of how demographic variables and group size predict time at the table (our proxy for engagement) for each exhibit software and study condition. We then turn to an examination of how experiences at the exhibit predict learning outcomes. The first set of analyses explore whether group interaction is associated with increased engagement (Evaluation Question 2). The second set of analyses explore whether increased engagement is associated with deeper understanding of key concepts of evolution (Evaluation Question 3).

In general, we find that larger groups, especially those who seem to know each other (or at least come and go together) spend longer at the exhibits, suggesting that the exhibit design encourages and builds upon positive group interactions. In addition, for DeepTree, time at the exhibit is associated with higher levels of agreement with evolution and common ancestry questions – important learning goals for the project – after controlling for educational level. Though our observational studies cannot make causal claims, these findings are consistent with the findings of the experimental learning research and together suggest that learning outcomes may be a direct result of exposure to the exhibit software.

Details of our analytic models and specific findings follow.

### Nested Models for Time Spent

As previously mentioned, the evaluation team has tried whenever possible to account for the fact that people experienced the exhibit as a part of informal groups. They experienced the exhibit together, sharing activities and conversation and, presumably, opportunities to learn. Because of this clustering, the experiences of individuals is not independent of one another, and ordinary least squares (OLS) multiple regression analysis would be inappropriate and would violate important statistical assumptions, as it would ignore the shared variability of people's experiences. Therefore, to predict people's time spent at the exhibit, hierarchical linear modeling using maximum likelihood estimation was used to account for these relationships using R package Multilevel (Bliese, 2012).

*DeepTree Video Study*

First, time spent at the exhibit was log transformed to normalize the data's distribution. Next, an unconditional model was fit predicting time spent from nothing but group averages plus individual error. Examining this model, we find that 96% of the variation in time spent is attributable to group membership (the intraclass correlation, or *ICC*). Based on hypotheses about what might affect dwell time, we explored several factors – presence in the group of child visitors of different ages (YNGCHILD (aged 5 or under), CHILD (6 to 12) and TEEN variables below), and the number of times (if any) that visitors encountered the "Relate glitch" software error that occurred during early evaluation testing during their time at the table (GLITCH). We define two group size variables: EXCLGRPSIZE represents the number of people in the group with whom the visitor mutually shared the majority of their time – this variable is centered at the median group size of 2. INGRPSIZEDIFF represents the difference between the modal group size for a person and EXCLGRPSIZE – that is, it's zero when the person only spent significant time at the table with people who were also in their mutual/ exclusive group, and positive when several people who were not part of the person's exclusive group also spent time at the exhibit. The full model is:

Level 1: $LOGTIME_{ij} = \beta_{00} + \beta_{10}GLITCH + \beta_{20}INGRPSIZEDIFF + r_{ij}$

Level 2: $\beta_{00} = \gamma_{00} + \gamma_{01}YNGCHILD + \gamma_{02}CHILD + \gamma_{03}TEEN + \gamma_{04}EXCLGRPSIZE + u_{0j}$

$\beta_{10} = \gamma_{10}$

$\beta_{20} = \gamma_{20}$

The grand mean log-time spent across all groups was 2.50 (319 seconds, or 5.3 minutes) (*SE* = .061). The effect of having a child or teen in the group on the group mean was not significant – groups with or without children present spent approximately the same amount of time at the exhibit. Each time the Relate Glitch occurred, people spent about 5% longer at the table ($\beta_{10}$ = .021, *SE* = .0089, *t* (92) = 2.38, *p* = .020) – it's clear from the videos that they spend time trying to figure out what the software is doing and trying to make it work. The relationship between group size and time at the table is somewhat complicated. In this full model, mutual group size doesn't predict time at the table, but the impact of additional people present for substantial time periods beyond the mutually shared group is negative, such that for each additional such person outside the mutually shared group, individuals spend only about 83% of the time ($\beta_{20}$ = –.083, *SE* = .0295, *t* (92) = –2.80, *p* = .0063).

Because presence of children of different ages had no impact on dwell time, we fit a model that only contains the GLITCH indicator and the two group size indicators:

Level 1: $LOGTIME_{ij} = \beta_{00} + \beta_{10}GLITCH + \beta_{20}INGRPSIZEDIFF + r_{ij}$

Level 2: $\beta_{00} = \gamma_{00} + \gamma_{04}EXCLGRPSIZE + u_{0j}$

$\beta_{10} = \gamma_{10}$

$\beta_{20} = \gamma_{20}$

In this model, grand mean log-time spent is 2.54 (347 seconds, 5.8 minutes) (*SE* = .046); each time visitors experience the Relate Glitch they spend 4% longer – about 14 seconds on average for the first occurrence ($\beta_{10}$ = .017, *SE* = .0086, *t* (92) = 1.99, *p* = .050). Now, the presence of additional mutually shared group members somewhat increases time at the exhibit, by about 18% for each additional group member within the range of the data, and this parameter is marginally significant ($\gamma_{04}$ = .072, *SE* = .0408, *t* (70) = 1.77, *p* = .082), and the presence of additional people outside the mutually shared group is associated with reduced time at the exhibit of 83% for each additional person ($\beta_{20}$ = –.080, *SE* = .0295, *t* (92) = –2.71, *p* = .0080). (See Table 20 for a summary of these results.)

Using a model comparison test to check whether a more parsimonious model is a better fit to the data, we find this last model is significantly better than an unconditional model in predicting dwell time, $\Delta\chi^2$ (3)= 16.69, *p* = .0008. We also tested a model that includes a quadratic term for group size, to test for non-linearity in the relationship between group size and dwell time. We reject this model because it is not a better fit to the data.

Thus, we find, after controlling for the added time visitors spent trying to figure out what was going on when the exhibit was experiencing the Relate software glitch, that larger mutual groups are associated with increased dwell time, but that additional people outside the mutually shared group are associated with reduced time at the exhibit. People seem to spend longer at the exhibit when they have familiar others to interact with.

*DeepTree Naturalistic Study*

In an attempt to replicate the findings from the DeepTree Video Study, hierarchical models were run for time spent at the exhibit in the Naturalistic study. We attempted to replicate the models run for the Video study; however we do not include predictors for GLITCHes, because the Relate software error was corrected before the naturalistic observation study, so it is expected that this effect would have disappeared.

As above, time spent at the exhibit was log transformed to normalize the data's distribution. Next, an unconditional model was fit predicting time spent from nothing but group averages plus individual error. The intraclass correlation for this model was somewhat less than for the video data (*ICC* = .80), suggesting somewhat more within group variability in time spent.

We tried to fit a model similar to the full model described above, including variables for the presence of children of different ages (YNGCHILD, CHILD and TEEN) as well as variables for group size. As above, we define the EXCLGRPSIZE variable to represent the number of people in the group with whom the visitor mutually shared the majority of their time (centered at the median group size of 2); and an INGRPSIZEDIFF variable to represent the difference between the modal group size for a person and EXCLGRPSIZE (i.e., centered at each person's EXCLGRPSIZE). The following model was fit to the data:

Level 1: $\text{LOGTIME}_{ij} = \beta_{00} + \beta_{10}\text{INGRPSIZEDIFF} + r_{ij}$

Level 2: $\beta_{00} = \gamma_{00} + \gamma_{01}\text{YNGCHILD} + \gamma_{02}\text{CHILD} + \gamma_{03}\text{TEEN} + \gamma_{04}\text{EXCLGRPSIZE} + u_{0j}$

$\beta_{10} = \gamma_{10}$

$\beta_{20} = \gamma_{20}$

The grand mean log-time spent across all groups, controlling for other variables (i.e., assuming a group of 2 adults) was 1.99 (98 seconds, or 1.63 minutes) (*SE* = .058). As noted previously, this is substantially less than for the video study. In general, groups with children of different ages spent less time at the exhibit than groups consisting of only adults. The effect of the presence of a young child on the group mean was significant, such that groups with young children spent an average of 54% of the time (average of 52 seconds) of those without young children (change in log(time) when a young child is present is $\gamma_{01}$= –.27, *SE* = .113, *t* (174) = –2.40, *p* = .017). The effects of having a target-age child or teenager in the group were slightly negative, but were not statistically significant, suggesting that such groups spent about as long at the exhibit as those consisting of just adults. (See Table 20 for a summary of these results.)

We also tested the impact of number of people around the exhibit on dwell time, controlling for the presence of different aged children. The coefficient for number of mutual table participants was significant, $\gamma_{04}$ = .106, *SE* = .044, *t* (174) = 2.40, *p* = .018. This suggests a 28% increase in time spent for each additional person in the mutual group – in general larger groups stay longer than smaller ones, within the limits of the data collected. However, this is moderated by an almost identical statistically significant *negative* impact of additional people present for substantial time periods *beyond* the mutually shared group, averaging 78% of time spent for each additional non-mutual person in the modal group. The coefficient for the group-centered difference in number of table participants (INGRPSIZEDIFF) was $\beta_{10}$ = –.107, *SE* = .034, *t* (133) = –3.10, *p* = .002.

Table 20: Final Nested Models Predicting Log-Time at DeepTree Exhibit

| | Video Study | | Naturalistic Study | |
|---|---|---|---|---|
| **Predictor** | **Estimate (SE)**<br>**$10^{Est}$ / *Ratio*** | **t** | **Estimate (SE)**<br>**$10^{Est}$ / *Ratio*** | **t** |
| Intercept (when EXCLGRPSIZE = 2) | 2.54 (.046)<br>347 sec | 54.93 *** | 1.99 (.058)<br>98 sec | 34.19*** |
| Mutual Group Size (EXCLGRPSIZE) | .072 (.041)<br>*1.18* | 1.77~ | .106 (.044)<br>*1.28* | 2.40* |
| Others at the Table beyond Mutual Group (INGRPSIZEDIFF) | –.080 (.030)<br>*0.83* | –2.71 ** | –.107 (.034)<br>*0.78* | –3.10** |
| Relate Software Bug (GLITCH) | .017 (.009)<br>*1.04* | 1.99 * | — | — |
| Presence of a Young child (YNGCHILD) | | | –.271 (.113)<br>*0.54* | –2.40 * |
| Presence of a Child (CHILD) | | | –.133 (.082)<br>*0.74* | –1.62 |
| Presence of a Teen (TEEN) | | | –.028 (.089)<br>*0.94* | –0.32 |
| Model Statistics | –2LogLikelihood = 106.25 (df=6) | $\Delta\chi^2$ (df=3) = 16.69<br>*p* = .0008 | –2LogLikelihood = 332.03 (df=8) | $\Delta\chi^2$ (df=5) = 28.69<br>*p* <.0001 |

Note: ~ p<.10; *p<.05; ** p<.01; *** p<.001
$\Delta\chi^2$ statistic compares this model with the unconditional model for the same data.

As above, we tested whether a more parsimonious model is a better fit to the data. A model comparison test suggests that the current model is significantly better than an unconditional model in predicting dwell time, $\Delta\chi^2$ (5)= 28.69, *p* < .0001. We also fit a model that included a quadratic term to test for non-linearity in the relationship between group size and dwell time, and found no statistically significant difference.

While the presence of young children is associated with reduced dwell time for the naturalistic study (but not the video study), as in the Video Study, larger familiar groups are associated with longer dwell times, while the presence of people beyond the mutual group tends to cancel this effect.

*BAT Video Study*

A similar analysis of the BAT video data used variables for age categories (YNGCHILD and CHILD; there were no TEENs in this study), group size (centered at the median size of 2), and whether or not to predict time at the exhibit (LOGTIME, transformed as above to account for the distributional shape).

Level 1: $LOGTIME_{ij} = \beta_{00} + r_{ij}$

Level 2: $\beta_{00} = \gamma_{00} + \gamma_{01}YNGCHILD + \gamma_{02}CHILD + \gamma_{03}EXCLGRPSIZE + u_{0j}$

The sample for this study was small – there were only 23 observations and 13 groups in the data set. Intraclass correlation (*ICC*) for the unconditional model is .88. The grand mean log-time spent across all groups, controlling for other variables (i.e., assuming a group of 2 adults) was 2.57 (372 seconds, or 6.21 minutes; $SE = .050$, $t\,(10) = 51.60$, $p < .0001$). The effects of group size and of having a young child in the group (< 6 years) were not statistically significant. The effect of having a target-age child (between 6 and 12 years) was statistically significant and positive ($\gamma_{02} = .536$, $SE = .095$, $t(9) = 5.65$, $p = .0003$) such that (the three) groups that included a target-age child spent 3.4 times as long (343%) at the exhibit on average than those that didn't (1286 seconds, 21.5 minutes). This finding is different from the DeepTree study where groups with children spent less time rather than more time, but it may be an artifact of the limited numbers in this sample – only 3 groups contained a target-age child. (See Table 21 for a summary of these results.)

We tested whether a more parsimonious model is a better fit to the data. A model comparison test suggests that the model that includes age characteristics is significantly better than an unconditional model in predicting dwell time, $\Delta\chi^2\,(3) = 17.99$, $p = .0004$.

*BAT Naturalistic Study*

For the BAT naturalistic data, we used variables for age categories (YNGCHILD, CHILD and TEEN) and group size (centered at the median size of 2) to predict time at the exhibit (LOGTIME, transformed as above to account for the distributional shape).

Level 1: $LOGTIME_{ij} = \beta_{00} + r_{ij}$

Level 2: $\beta_{00} = \gamma_{00} + \gamma_{01}YNGCHILD + \gamma_{02}CHILD + \gamma_{03}TEEN + \gamma_{04}EXCLGRPSIZE + u_{0}$

There were 155 observations and 95 groups in the data set. Intraclass correlation for the unconditional model is .96 — suggesting little variability in time spent at the exhibit among people who were part of the same mutual group. The grand mean log time spent across all groups, controlling for other variables (i.e., assuming a group of 2 adults) was 1.93 (86 seconds, or 1.43 minutes; $SE = .094$ $t(90) = 20.51$, $p < .0001$).

Table 21: Final Nested Models Predicting Log-Time at BAT Exhibit

| Predictor | Video Study: Estimate (SE) $10^{Est}$ / *Ratio* | Video Study: t | Naturalistic Study: Estimate (SE) $10^{Est}$ / *Ratio* | Naturalistic Study: t |
|---|---|---|---|---|
| Intercept (when EXCLGRPSIZE = 2) | 2.57 (.050)<br>372 sec | 51.60*** | 2.00 (.058)<br>100 sec | 34.31 *** |
| Mutual Group Size (EXCLGRPSIZE) | .013 (.047)<br>*1.03* | 0.29 | .115 (.057)<br>*1.30* | 2.04* |
| Presence of a Young child (YNGCHILD) | .049 (.107)<br>*1.12* | 0.45 | | |
| Presence of a Child (CHILD) | .536 (.095)<br>*3.43* | 5.65*** | | |
| Model Statistics | –2LogLikelihood = –30.12 (df=6) | $\Delta\chi^2$ (df=3) = 17.99<br>*p* = .0004 | –2LogLikelihood = 91.75 (df=5) | $\Delta\chi^2$ (df=1) = 4.08<br>*p* = .043 |

Note: ~ p<.10; *p<.05; ** p<.01; *** p<.001
$\Delta\chi^2$ statistic compares this model with the unconditional model for the same data.

As for the DeepTree Naturalistic study, for the BAT Naturalistic study, increased numbers of people around the exhibit table has a generally positive relationship with dwell time. The coefficient for number of mutual table participants was significant, $\gamma_{04} = .121$, *SE* = .058, *t* (90) = 2.09, *p* = .039. This suggests a 32% increase in time spent for each additional person in the mutual group – in general larger groups stay longer than smaller ones, within the limits of the data collected. Presence of young children tended to decrease time at the table ($\gamma_{01} = -.096$, ratio = .80, *SE* = .138, *t*(90) = –0.69, *p*=.49); presence of children tended to increase time ($\gamma_{02} = .176$, ratio = 1.50, *SE* = .109, *t*(90) = 1.61, *p*=.11); and presence of teens tended to increase time ($\gamma_{03} = .025$, ratio = 1.06, *SE* = .137, *t*(90) = 0.18, *p*=.86), but none of these differences are statistically significant. Because presence of children of different ages had no statistically significant impact on dwell time, we fit a model that only contains the centered mutual group size indicator:

Level 1: $LOGTIME_{ij} = \beta_{00} + r_{ij}$

Level 2: $\beta_{00} = \gamma_{00} + \gamma_{04}EXCLGRPSIZE + u_0$

In this model, grand mean log-time spent is 2.00 (100 seconds, 1.67 minutes) (*SE* = .058). The presence of additional mutually shared group members increases time at the exhibit by about 30% for each additional group member within the range of the data ($\gamma_{04} = .115$, *SE* = .057, *t* (93) = 2.03, *p* = .045). (See Table 21 for a summary of these results.)

Using a model comparison test to check whether a more parsimonious model is a better fit to the data, we find this last model is significantly better than an unconditional model in predicting dwell time, $\Delta\chi^2$ (1)= 4.08, *p* = .043. We also tested a model that includes a quadratic term for group size, to test for non-linearity in the relationship between group size and dwell time. We reject this model because it is not a better fit to the data.

There are interesting similarities (and a few differences) in the dwell time results for the DeepTree and BAT studies:

- Mean dwell time for groups of two (2) without other distinguishing characteristics is about 100 seconds in the naturalistic conditions, and about 6 minutes in the video condition in both studies.
- For both exhibits, group characteristics substantially predict dwell time (they have high intraclass correlations). In the DeepTree exhibit, this was somewhat more true in the video than the naturalistic condition; in the BAT study, it was somewhat more true in the naturalistic than the video study. Mostly, people came and went as groups, but there was more variation in the DeepTree naturalistic study than the others when visitors were freer to come and go and didn't have to formally enter and exit the cordoned research area.
- In the naturalistic studies of both exhibits, additional people around the table who are part of visitors' mutual groups are associated with about a 30% increase in time at the table, within the extent of the data. In the video studies, there is an increase in dwell time with larger mutually shared groups, but it's only an 18% increase in the DeepTree study, and just 3% in the BAT study, and marginally statistically significant, if at all.
- In both study conditions for the DeepTree software (but not BAT), additional people at the table who are not part of visitors' mutual groups are associated with a decline in dwell time to about 80% of the time when such people are not present. Thus, the addition of "strangers" (i.e., those who don't come and go together) reduces dwell time where larger mutual groups tend to spend longer.
- Presence of children in the group is associated with reduced time at the table for the DeepTree naturalistic condition, though only the presence of young children gives a statistically significant difference. In the DeepTree video study, presence of young children is associated with reduced time at the table, while presence of target-age children and teens is associated with slightly increased time at the table, though none of these results are statistically significant.
- By contrast, for BAT presence of children tended to increase dwell time, however these associations are not statistically significant except for the dramatic case of target-age children at the BAT video study (though this represents only 3 groups).

Overall, these findings provide evidence that the *Life on Earth* exhibit – both DeepTree and BAT – was successful at engaging groups for substantial times with exhibit activities, and that larger groups were engaged for longer on average than smaller ones, suggesting the importance of social interaction around the exhibit table.

## Nested Models for Learning Outcomes – DeepTree/ FloTree

For each of the learning composite scores (common ancestry and evolution), an organization effects model was fit to people nested in groups. Fitting unconditional models for both common ancestry and evolution showed significant variability in group scores, on average. The *ICC* for Common Ancestry scores was .35 and for Evolution scores it was .11. This suggests that for both variables – more for Common Ancestry than Evolution – some variability in scores depends on group factors, and some depends on individual variation within the group. Therefore, conditional models adding predictors at Level 1 (person-level) and Level 2 (group level) were appropriate. At Level 1, we tested whether various measures of engagement with the exhibit — whether people experienced the Relate function (RELATE) or the FloTree function (FLOTREE); how much Biology talk occurred while they were at the table, log transformed to normalize the distributional shape of the variable (LOGBIOTALK); overall time spent (log transformed and centered at the mean (LOGTIME), so that other variables can be interpreted at the average time spent) — as well as whether people's education predicted these composite scores (ED, centered at the median (4-year college education) so other coefficients can be interpreted for the average visitor). Below, the results of the conditional models are presented separately for each composite score.

*Common Ancestry Results*

In this model, 123 people in 55 different mutually exclusive groups were used for analysis. The model was fit as follows:

Level 1: $COMANC_{ij} = \beta_{00} + \beta_{10}LOGBIOTALK + \beta_{20}FLOTREE + \beta_{30}RELATE + \beta_{40}ED + \beta_{50}LOGTIME + r_{ij}$

Level 2: $\beta_{00} = \gamma_{00} + u_{0j}$

$\beta_{10} = \gamma_{10}$

$\beta_{20} = \gamma_{20}$

$\beta_{30} = \gamma_{30}$

$\beta_{40} = \gamma_{40}$

$\beta_{50} = \gamma_{50}$

The grand mean group score for common ancestry across all individuals controlling for group membership and other variables was 3.75 (*SE* = .24). This is interpreted directly — average agreement for the common ancestry composite for a college educated visitor who spent an average amount of time at the exhibit, and didn't engage in biology talk or use Relate or FloTree, was 3.75, somewhat above a neutral rating of 3.00.

The effects of biology talk, finding Relate, and finding FloTree were non-significant. Visitors' agreement with statements about common ancestry did not depend on whether they used the Relate or FloTree features, or the extent of their biology talk, after controlling for other variables.

Not surprisingly, the effect of education on an individual's agreement with statements about common ancestry was marginally significant ($\gamma_{40}$ = .062, *SE* = .033*, t*(63) = 1.91, *p* = .060), such that each additional increment of schooling (e.g., some high school to high school grad; or 2 year college to 4 year college) gave a small (.062) additional increment in common ancestry scores. The difference between an elementary school student and a college grad is predicted to be .37 points – a figure in line with the learning research which found the mean of the composite score for its elementary grade participants to be around 3, with significant differences by age.

The effect of people's exhibit dwell time on common ancestry scores was statistically significant controlling for these other factors ($\gamma_{50}$ = 1.08, *SE* = .409*, t*(63) = 2.64, *p* = .010). Because dwell time is log-transformed in the model, interpretation is not as direct as for other variables. Transforming back within the range of variation of the variable, this suggests that a doubling of time at the exhibit is associated with a .33 point increase in common ancestry scores, all else being equal. This does not imply causality in our study – it is not clear whether time at the exhibit "boosted" scores, or whether those who had a deeper understanding of common ancestry stayed longer at the exhibit. However, the learning research's experimental design *did* show a causal link between exposure to the *Life on Earth* exhibit and acceptance of common ancestry, so a similar effect may be occurring among visitors in the field. (See Table 22 for a summary of these results.)

We tested whether a more parsimonious model is a better fit to the data. A model comparison test suggests that the current model is significantly better than an unconditional model in predicting common ancestry scores, $\Delta\chi^2$ (5)= 15.99, *p* = .007.

*Evolution Results*

The same model as described above was fit for the composite level of agreement on statements about evolution, with similar results. One hundred and twenty-two (122) individuals in 56 groups were used for this analysis. The grand mean group score for understanding of evolution across all individuals was 4.12 (*SE* = .17), indicating moderate to strong levels of agreement on average.

Table 22: Hierarchical Models Predicting Learning Outcomes at DeepTree Exhibit, Video Study

| | Common Ancestry | | Evolution | |
|---|---|---|---|---|
| **Predictor** | **Estimate (SE)** | **t** | **Estimate (SE)** | **t** |
| Intercept | 3.74 (.24) | 15.44 *** | 4.12 (.172) | 24.01*** |
| Biology Talk (BIOTALK) | –.095 (.218) | –0.44 | .242 (.153) | 1.58 |
| Use of FloTree (FLOTREE) | –.014 (.240) | –0.06 | –.069 (.168) | –0.42 |
| Use of Relate (RELATE) | .350 (.255) | 1.37 | .103 (.177) | 0.58 |
| Education Level (ED) | .062 (.033) | 1.91~ | .066 (.026) | 2.58 * |
| Time at Exhibit (LOGTIME) | 1.08 (.409) | 2.64 * | .545(.293) | 1.86~ |
| Model Statistics | –2LogLikelihood = 333.30 (df=8) | $\Delta\chi^2$ (df=5) = 15.99 *p* = .007 | –2LogLikelihood = 263.62 (df=8) | $\Delta\chi^2$ (df=5) = 15.17 *p* =.010 |

Note: ~ p<.10; *p<.05; ** p<.01; *** p<.001

The effects of finding the Relate or FloTree functions, or of extent of biology talk, on levels of agreement about these statements about evolution were non-significant.

The effect of education on an individual's learning score was significant ($\gamma_{40}$ = .066, *SE* = .026*, t*(61) = 2.58, *p* = .012), such that higher educated individuals gave higher agreement ratings for the evolution questions, on average, but again only very slightly, with a .39 increment for the elementary student to college graduate range described above. Again, these results align with those found in the project's learning research, where there was a strong effect for age.

The effect of people's exhibit dwell time on evolution scores was marginally statistically significant ($\gamma_{50}$ = .545, *SE* = .293*, t*(61) = 1.86, *p* = .068). As above, we interpret this effect by transforming the coefficient back within the range of variation of the variable. This suggests that a doubling of time at the exhibit is associated with a .16 point increase in evolution scores, all else being equal. As above, this does not imply causality, however – it is not clear whether time at the exhibit "boosted" scores slightly, or whether those who had a deeper understanding of evolution stayed longer at the exhibit. This marginal result may also align with the learning research which found no significant results by experimental condition. (See Table 22 for a summary of these results.)

This model was a significant improvement in fit over an unconditional model in predicting evolution learning scores, $\Delta\chi^2$ (5)= 15.17, *p* = .010.

These analyses suggest that with the DeepTree software, once group characteristics and visitors' prior level of education are controlled for, greater time at the exhibit is associated with a small but statistically significant increase in common ancestry scores, and marginally associated with a small increase in evolution scores. Although causality cannot be attributed solely from these findings – those who know more about common ancestry and evolution may stay longer at the exhibit; or staying longer at the exhibit may lead to these increased scores – their alignment with the results from the learning research which did have an experimental design suggest that these increased scores may be a direct result of exposure to the exhibit software. This is an important finding.

## Nested Models for Learning Outcomes – Build-a-Tree (BAT)

*Common Ancestry Results*

In this model, 18 people in 11 different mutually exclusive groups were used for analysis, with the time log-transformed and centered, and education centered at the median college grad level, as above. Both because the Build-a-Tree software doesn't include FloTree or Relate, and because we didn't have table log data for Build-a-Tree, our model is simpler than for DeepTree:

Level 1: $COMANC_{ij} = \beta_{00} + \beta_{10}LOGBIOTALK + \beta_{20}ED + \beta_{30}LOGTIME + r_{ij}$

Level 2: $\beta_{00} = \gamma_{00} + u_{0j}$

$\beta_{10} = \gamma_{10}$

$\beta_{20} = \gamma_{20}$

$\beta_{30} = \gamma_{30}$

For BAT, in the complete model, the effect of biology talk and time at the exhibit were not statistically significant and in a model that includes them, the effect of education was only marginally significant, therefore we report on the simpler model that only includes level of education. Intraclass correlation for the unconditional model is *ICC* = .42. The grand mean group score for common ancestry across all individuals controlling for group membership and education level was 3.73 ($SE$ = .25, $t(10)$ = 15.00, $p$ <.0001). This is interpreted directly — average agreement for the common ancestry composite for a college educated visitor was 3.73, somewhat above a neutral rating of 3.00.

The effect of education on an individual's agreement with statements about common ancestry using the BAT software was larger than in the DeepTree study, and statistically significant ($\gamma_{20}$ = .227, $SE$ = .092, $t(63)$ = 2.48, $p$ = .048), such that each additional increment of schooling (e.g., some high school to high school grad; or 2 year college to 4 year college) gave a moderate (.23) additional increment in common ancestry scores. (The difference between an elementary school student and a college grad is predicted to be 1.37 points on a 5 point scale.)

We tested whether a more parsimonious model is a better fit to the data. A model comparison test suggests that the model including educational level is significantly better than an unconditional model in predicting common ancestry scores, $\chi^2$ (1)= 5.57, $p$ = .018.

*Evolution Results*

Although we initially fit this model with 18 people in 11 different mutually exclusive groups, the intraclass correlation is essentially zero, suggesting that group membership doesn't affect people's evolution scores. Overall average evolution score in this group is also very high, at 4.52. Fitting non-hierarchical OLS regression models, we find that neither educational level, time at the table, extent of biology talk, or even common ancestry score has a statistically significant effect on evolution scores for the BAT exhibit.

## Discussion/ Conclusions

The *Life on Earth* exhibit software was successful at engaging a wide range of visitors with a variety of activities to learn about diversity of species, common ancestry, and evolutionary processes. Visitors spent about 1.7 minutes on average in the naturalistic conditions for both DeepTree and BAT, and between 5.8 and 6.2 minutes on average in the video condition, controlling for group characteristics. These are substantial levels of engagement for museum exhibits. For the DeepTree video study, a majority of visitors (59%) engaged in 4 or more of the 7 major exhibit activities, and 38% engaged in 5 or more. Again, these are substantial levels of engagement with content. Although our attempt to create a reliable multi-dimensional measure of exhibit engagement through confirmatory factor analysis (CFA) was unsuccessful, it did show that dwell time was the most reliable predictor of a hypothesized underlying broader latent engagement factor, supporting our use of this variable both as an outcome and predictor variable in our hierarchical modeling.

Although some people had difficulty figuring out how the software was supposed to work (which was exacerbated by a small number of software bugs during the DeepTree video study), for the most part they found ways to engage with the software in both intended and novel ways. Novel behaviors included:

- Encyclopedic scanning of species (via pictures and text) across the canopy, often done by parents with younger children;
- Use of reel or top item navigation to find a species at the canopy, then manual navigation to back up from the canopy to a common ancestor, and then again forward down a branch to a related descendant;
- Group attempts to "kill off" one branch in FloTree, observed both with young adults and groups of kids; and
- Use of an iPad by one teenager to look up information about specific species observed on the table, while his parents explored the exhibit software, among others.

Visitors say they found both the DeepTree and BAT exhibits interesting and enjoyable, and that they would encourage their friends to visit it. Observations show visitors verbally expressing enjoyment with the exhibit in nearly half of DeepTree visitors' experiences, and over 70% of BAT visitors' experiences. While some visitors expressed frustration or displeasure with the exhibit (including a substantial minority of older adults who said the zooming effect made them dizzy), such verbal expressions were only observed in 25% or fewer of visitors' experiences.

The exhibit was designed for groups of visitors to work together with a common focus around the table at the same time, rather than working in parallel, and was successful in creating the conditions in which various social interactions among visitors were common. Groups of visitors often navigated simultaneously – usually successfully but with some conflicts observed especially among kids who wanted to control the software and were frustrated by its responses to multiple touches. Still, visitors often took turns controlling the software, or used verbal (social negotiation) or physical (pointing/ indicating) means to negotiate what to do next.

These negotiations around how to use the software together were accompanied by frequent talk about biological content – majorities of visitors in the naturalistic studies (50% DeepTree; 65% BAT) and substantial majorities in the video studies (over 80%) experienced some talk about biology content while at the exhibit table. This included questions about whether species were related, comments about characteristics of varied species and/ or reading information displayed in the exhibit, and hypotheses about what was going on in the FloTree "Experiment," among others.

In addition, a substantial minority of visitors were able to connect their exhibit experiences to other experiences at the museum and elsewhere in their lives – e.g., learning from classes they had taken, or diseases they or their friends had experienced – with about 15% in all study and software conditions expressing one or more such connections during their time at the exhibit. This suggests some integration and meaning-making of the *Life on Earth* experience into visitors' existing conceptual and experiential frameworks.

The social nature of the exhibit design was effective. Visitor groups represented a wide variety of age configurations with more fluid overlap in the naturalistic studies than in the video studies, as expected because of the way we controlled access to the video study. While group sizes varied from 1 to 7, median group size was 2 or 3 in each software and study condition. Hierarchical analyses show that larger mutual groups at the exhibit were associated with increased dwell time for both DeepTree and BAT, with statistically significant differences in both naturalistic conditions, and marginally significant differences for the DeepTree (but not the BAT) video condition. In the DeepTree study (but not BAT), the effect of a larger group is positive if the increased numbers mutually share their time at the exhibit – presumably because they came together/ form a family or friendship group – but the effect is cancelled or becomes negative when there are additional people at the table who are *not* part of a mutually shared group. Thus, for example, our model predicts a group of three who came together would spend 28% more time than a group of two who came together, but a group of three consisting of two who came together and an additional person would spend about 20% less time. Perhaps people reduce time they otherwise would spend at the exhibit when they feel others are waiting; or they may feel less comfortable exploring with strangers than they do with family and friends.

This effect of additional people on dwell time is important both as an outcome in its own right – indicating that the software was successful in creating an environment in which groups of people could engage with the exhibit content together – and because of the association with learning outcomes summarized below.

The impact on dwell time of the presence of children of various ages was complicated, but minor. Groups containing children or especially younger children tended to spend less time in the DeepTree naturalistic study than did groups containing only adults or those with teens, though these differences are only statistically significant for the youngest children. DeepTree was designed for ages 10 and up, and this dwell time evidence suggests that, as intended, some of the content of DeepTree may be more interesting and engaging for teen and adult learners than for younger children.

The presence of 6-12 year old children (and to a lesser extent, teens) was associated with *increased* time in the BAT studies, though the difference is not statistically significant in the naturalistic study and, though statistically significant in the video study, is based on just 3 groups so may be partially an artifact. Still, it seems the BAT software tended to engage groups with children somewhat more than groups with just adults. In general, it seems the *Life on Earth* software works as well for groups with children as those without.

The exhibit was also successful in engaging visitors with biological content and, as noted above, large proportions of visitors engaged in biology talk while around the exhibit. Substantial fractions of survey respondents correctly described the exhibit as about evolution or relatedness or (for DeepTree) the tree of life. Hierarchical analyses show that, after controlling for visitors' level of education, dwell time at the DeepTree software (but not BAT) was associated with small, but significantly higher scores on the common ancestry and evolution scales used in the learning research. Though not causal, this is an important finding, directly linking engagement with the DeepTree software with higher scores on important learning outcomes/ goals. This result did not vary by whether or not visitors experienced important features of the DeepTree software – Relate or FloTree – or by the extent of biological talk they engaged in while at the exhibit. This suggests that the associations described do not depend on specific behaviors, but are more generally true across a range of experiences with the exhibit.

A similar result was found with the project's own learning research studies – a controlled experiment that used the same questions about common ancestry and evolution. In that study, youth who experienced the DeepTree exhibit were more likely to agree with the common ancestry questions. In the learning research studies, there was also a positive association between use of the Relate function and increased agreement with common descent. This result was stronger for the younger, 8-11 year-olds, which may be why it was not replicated in the current evaluation, which included adult as well as youth participants in the sample.

The non-causal, observational association that we found between dwell time at the DeepTree exhibit and ratings of agreement with learning outcomes, along with the causal link found by the learning research, together suggest that exposure to the exhibit software in the naturalistic museum context may increase scores on important learning outcomes.

Thus, overall, we can conclude that the *Life on Earth* exhibit was successful at engaging a wide range of visitors with important scientific content, encouraging group interaction, talk, and play, and perhaps even leading to deeper understanding of evolution and common ancestry.

## Bibliography/ References


Apple, Inc. (2009). Final Cut Pro 7 (Version 7): Apple, Inc.

Bliese, Paul. (2012). Multilevel Functions (Version 2.4).

Block, Florian, Horn, Michael S., Caldwell Phillips, Brenda, Diamond, Judy, Evans, E. Margaret, & Shen, Chia. (2012). The DeepTree Exhibit: Visualizing the Tree of Life to Facilitate Informal Learning. *IEEE Transactions on Visualization and Computer Graphics (TVCG), 18*(12), 2789-2798.

Borun, M., Chambers, M., & Cleghorn, A. (1996). Families are learning in science museums. *Curator, 39*(2), 123-138.

Butts, Carter T. (2013). Tools for Social Network Analysis (Version 2.3-1). Retrieved from http://www.statnet.org

Byrt, T., Bishop, J., & Carlin, J. B. (1993). Bias, prevalence and Kappa. *Journal of Clinical Epidemiology, 46*(5), 423-429.

Chua, K.C., Qin, Y.Q., Block, Florian, Phillips, Brenda, Diamond, Judy, Evans, E. Margaret, . . . Shen, Chia. (In Press 2013). *FloTree: A Multi-touch Interactive Simulation of Evolutionary Processes.* Paper presented at the Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces.

Davis, P., Horn, Michael S., Schrementi, L., Block, Florian, Phillips, Brenda, Evans, E. Margaret, . . . Shen, Chia. (2013). *Going Deep: Supporting collaborative exploration of evolution in natural history museums*. Paper presented at the International Conference on Computer Supported Collaborative Learning (CSCL'13), Madison, Wisconsin.

Evans, E. Margaret, Horn, Michael S., Phillips, Brenda C, Block, Florian, Diamond, Judy, & Shen, Chia. (2013). Life on Earth: Learning Research Report: University of Michigan.

Evans, E. Margaret, Phillips, Brenda C, Horn, Michael S., Block, Florian, Diamond, Judy, & Shen, Chia. (April, 2013). *Active Prolonged Engagement: When Does it Become Active Prolonged "Learning"?* Paper presented at the Biennial Meeting of the Society for Research in Child Development, Seattle, WA.

Evans, E. Margaret, Spiegel, Amy, Gram, W., Frazier, B. F., Tare, M., Thompson, S., & Diamond, Judy. (2010). A conceptual guide to natural history museum visitors' understanding of evolution. *Journal of Research in Science Teaching, 47*, 326-353. doi: 10.1002/tea.20337

Filemaker Inc. (2011). Filemaker Pro (Version 9): Filemaker, Inc.

Filemaker Inc. (2013). Filemaker Go for iPad. Santa Clara, CA.

Finzer, William. (2005). Fathom™ Dynamic Data™ Software (Version 2.0). Emeryville, CA: Key Curriculum Press.

Hallgren, Kevin A. (2012). Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutorials in Quantitative Methods for Psychology, 8*(1), 23-34.

Harris, A., Rick, J., Bonnett, V., Yuill, N., Fleck, R., Marshall, P., & Rogers, Y. (2009). *Around the table: Are multi-touch surfaces better than single-touch for children's collaborative interactions?* Paper presented at the Proceedings of the 9th International Conference on Computer Supported Learning (CSCL '09).

Higgins, S., Mercier, E., Burd, E., & Joyce-Gibbons, A. (2012). Multi-touch tables and collaborative learning. *British Journal of Educational Technology, 43*(6), 1041-1054.

Horn, Michael S., Leong, Z.A., Block, Florian, Diamond, Judy, Evans, E. Margaret, Phillips, Brenda, & Shen, Chia. (2012). *Of BATs and APEs: An interactive tabletop game for* *natural history museums*. Paper presented at the Proceedings of the ACM Conference on Human Factors in Computing Systems CHI'12.

Hornecker, E. (2008). *"I don't understand it either, but it is cool" Visitor Interactions with a Multi-Touch Table in a Museum*. Paper presented at the Proceedings of IEEE Tabletop 2008.

Humphrey, Thomas, Gutwill, Joshua P., & The Exploratorium APE Team. (2005). *Fostering Active Prolonged Engagement: The Art of Creating APE Exhibits*. San Francisco, CA: The Exploratorium.

IBM. (2011). SPSS Statistics (Version 20).

Jacucci, G., Morrison, A., Richard, G.T., Kleimola, J., Peltonen, P., Parisi, L., & Laitinen, T. (2010). *Worlds of information: designing for engagement at a public multi-touch display*. Paper presented at the ACM CHI '10: Proceedings of the 28th international conference on Human factors in computing systems.

Jöreskog, Karl, & Sörbom, Dag. (2012). LISREL (Version 8.80). Lincolnwood, IL: Scientific Software International.

Lindemann-Matthies, P., & Kramer, T. (2006). The influence of an interactive education approach on visitors' learning in a Swiss zoo. *Science Education, 90*(2), 296-315. doi: 10.1002/sce.20127

Microsoft. (2011). Excel for Macintosh (Version 14.3.6).

QSR International. (1999-2011). NVivo 9 (Version 9.1): QSR International Pty Ltd. Retrieved from www.qsrinternational.com

R Development Core Team. (2012). R: A language and environment for statistical computing (Version 2.15.0). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Raudenbush, Stephen W, Bryk, Anthony S., & Congdon, Richard. (2011). HLM for Windows (Version 7). Lincolnwood, IL: Scientific Software International.

Red Giant, LLC. (2012). PluralEyes 2 (Version 2).

Rick, J., Rogers, Y., Haig, C., & Yuill, N. (2009). Learning by Doing with Shareable Interfaces. *Children, Youth and Environments, 19*(1), 320-341.

Serrell and Associates. (2009). Summative Evaluation: Beautiful Science: Ideas that Changed the World.

Serrell, B. (1997). Paying attention: The duration and allocation of visitors' time in museum exhibitions. *Curator, 40*(2), 108-125.

Spiegel, Amy, Evans, E. Margaret, Frazier, B. F., Hazel, A., Tare, M., Gram, W., & Diamond, Judy. (2012). Changing museum visitors' concepts of evolution. *Evolution: Education and Outreach, 5*, 43-61. doi: 10.1007/s12052-012-0399-9

vom Lehn, D., & Heath, C. (2005). Accounting for new technology in museum exhibitions. *International Journal of Arts Management, 7*(3), 11-21.

Wasserman, Stanley, & Faust, Katherine. (1994). *Social Network Analysis: Methods and Applications*. Cambridge, UK: Cambridge University Press.

# Life on Earth Evaluation — Appendices

## Appendices Table of Contents

**Life on Earth Evaluation — Appendices** .......... A-i
**Appendices Table of Contents** .......... A-i
**Table of Appendix Tables and Figures** .......... A-ii
**Study Signage** .......... A-1
**Consent Text** .......... A-2
**Demographic Information** .......... A-3
DeepTree Video Participants .......... A-3
DeepTree Naturalistic Participants .......... A-5
BAT Video Participants .......... A-6
BAT Naturalistic Participants .......... A-8
**Observation Protocol Software** .......... A-10
**Observation Protocol Codebook** .......... A-11
**Observation Protocol Inter-Rater Reliability** .......... A-15
**Dwell Time Details** .......... A-17
By Age Group .......... A-17
By Group Size .......... A-18
**Event Proportions by Intervals** .......... A-20
DeepTree Event Proportions .......... A-20
BAT Event Proportions .......... A-22
**Table Log Summary Measures Definitions** .......... A-24
**Summary Statistics for Table Log Measures** .......... A-30
**Confirmatory Factor Analysis of Exhibit Engagement** .......... A-34
**Semantic Log Visualization Guide** .......... A-37
**Survey Text** .......... A-38
**Survey Results** .......... A-41
DeepTree Survey Results .......... A-41
BAT Survey Results .......... A-43

## Table of Appendix Tables and Figures

Table 23: DeepTree Video Observation Sample by Age Category and Gender ..... A-3
Table 24: DeepTree Video Study Survey Sample by Self-reported Age Category and Gender ..... A-3
Table 25: Self-identified Race/Ethnicity of DeepTree Video Observation Survey Sample ..... A-4
Table 26: DeepTree Video Participant-reported Education by Age Category ..... A-4
Table 27: DeepTree Naturalistic Observation Sample by Age Category and Gender ..... A-5
Table 28: DeepTree Naturalistic Study Survey Sample by Self-reported Age Category and Gender ..... A-5
Table 29: Self-identified Race/Ethnicity of DeepTree Naturalistic Survey Sample ..... A-5
Table 30: DeepTree Naturalistic Participant-reported Education by Age Category ..... A-6
Table 31: BAT Video Study Sample by Age Category and Gender ..... A-6
Table 32: BAT Video Study Survey Sample by Self-reported Age Category and Gender ..... A-6
Table 33: BAT Video study Self-identified Race and Ethnicity of Survey Sample ..... A-7
Table 34: BAT Video Study Participant-reported education by Age Category ..... A-7
Table 35: BAT Naturalistic Study Sample by Age Category and Gender ..... A-8
Table 36: BAT Naturalistic Study Survey Sample by Self-reported Age Category and Gender ..... A-8
Table 37: BAT Naturalistic Study Self-identified Race and Ethnicity of Survey Sample ..... A-8
Table 38: BAT Naturalistic Study Participant-reported education by Age Category ..... A-9
Table 39: Time Spent at Table by Age Category for DeepTree Video study ..... A-17
Table 40: Time Spent at Table by Age Category for DeepTree Naturalistic study ..... A-17
Table 41: Time Spent at Table by Age Category for BAT Video study ..... A-17
Table 42: Time Spent at Table by Age Category for BAT Naturalistic study ..... A-18
Table 43: Modal Number of People at Exhibit vs. Dwell Time for DeepTree Video Study ..... A-18
Table 44: Number of People at Exhibit vs. Dwell Time for DeepTree Naturalistic study ..... A-18
Table 45: Number of People at Exhibit vs. Dwell Time for BAT Video Study ..... A-19
Table 46: Number of People at Exhibit vs. Dwell Time for BAT Naturalistic Study ..... A-19
Table 47: Proportion of Engagement Behaviors (SD) over Total Available Time Intervals by Group Size for DeepTree Video Study ..... A-20
Table 48: Proportion of Engagement Behaviors over Total Available Intervals by Group Size for DeepTree Naturalistic Study ..... A-21
Table 49: Proportion of Engagement Behaviors over Total Available Intervals by Group Size for BAT Video Study ..... A-22
Table 50: Proportion of Engagement Behaviors over Total Available Intervals by Group Size BAT Naturalistic Study ..... A-23
Table 51: General Table Activity Descriptives ..... A-30
Table 52: Navigation Activity Descriptives ..... A-30
Table 53: Inspection Activity Descriptives ..... A-31
Table 54: Trait Activity Descriptives ..... A-31
Table 55: Relate Function Descriptives ..... A-32
Table 56: Training Tree Activity Descriptives ..... A-32
Table 57: FloTree Descriptives ..... A-33
Table 58: List of Indicators for Confirmatory Factor Analysis ..... A-34
Table 59: LOE Indicator Reliabilities (Item Reliability Estimates) ..... A-35
Table 60: Indicator-factor Correlation (Validity Estimates) ..... A-35
Table 61: Self-reported ratings of exhibit experience, DeepTree Video Study ..... A-41
Table 62: Levels of Agreement with Survey Questions about Common Ancestry of Various Species. DeepTree Video Study ..... A-41
Table 63: Levels of Agreement with Survey Questions about Evolution. DeepTree Video Study ..... A-41
Table 64: Self-reported ratings of exhibit experience, DeepTree Naturalistic Study ..... A-42
Table 65: Levels of Agreement with Survey Questions about Common Ancestry of Various Species. DeepTree Naturalistic Study ..... A-42

Table 66: Levels of Agreement with Survey Questions about Evolution. DeepTree Naturalistic Study ........ A-42
Table 67: Self-reported ratings of exhibit experience, BAT Video Study ........ A-43
Table 68: Levels of Agreement with Survey Questions about Common Ancestry of Various Species. BAT Video Study ........ A-43
Table 69: Levels of Agreement with Survey Questions about Evolution. BAT Video Study ........ A-43
Table 70: Self-reported ratings of exhibit experience, BAT Naturalistic Study ........ A-44
Table 71: Levels of Agreement with Survey Questions about Common Ancestry of Various Species. BAT Naturalistic Study ........ A-44
Table 72: Levels of Agreement with Survey Questions about Evolution. BAT Naturalistic Study ........ A-44

## Appendix Figures

Figure 20: Screen Shot of Observation Protocol Database ........ A-10
Figure 21: CFA – Table Engagement: Factor Loadings ........ A-35

## Study Signage

In order to inform museum visitors that evaluation research was happening at the museum, the following signs were created and posted

***Sign 1 - Museum entrance – On stanchion toppers***

Help Improve Our *Life on Earth* Exhibit
You may be observed or videotaped in some parts of the museum today.
Signs will be posted in the research areas.

***Sign 2 – Video Study Entrance –***
***Area cordoned off, access requires written informed consent***

RESEARCH IN PROGRESS
Help Us Improve Our *Life on Earth* Exhibit
We are videotaping and surveying people using this exhibit in order to make it better.
If you do not want to be videotaped, please visit this area later.
Video will be used for research and not for commercial or broadcast purposes.
Video may be shown at education or museum conferences to inform our colleagues.

***Sign 3 – Naturalistic Study – not cordoned of***

RESEARCH IN PROGRESS
Help Improve Our *Life on Earth* Exhibit
We are observing people using this exhibit in order to make it better.
If you do not want to participate, please visit this area later.
Observations are anonymous. No identifying information is collected.
Results may be shared at education or museum conferences to inform our colleagues.

## Consent Text

TERC Harvard School of Engineering and Applied Science

# *Life on Earth*
# Evaluation Research Information Sheet & Consent

**What is *Life on Earth*?** *Life on Earth* is an interactive touch table that tries to help people learn about different plants, animals and other species on earth and how they are related.
**Who is doing the evaluation?** The California Academy of Sciences is helping to test the *Life on Earth* touch table exhibit. The exhibit is being made and tested by researchers at Harvard, the University of Michigan, the University of Nebraska, Northwestern University (Evanston, IL), and TERC (Cambridge, MA). Funds for the project come from the National Science Foundation (NSF). Dr. Chia Shen at Harvard leads the project. Dr. James Hammerman at TERC leads the evaluation.
**Why are we doing the evaluation?** We are testing the exhibit to find out how well it works, and what people learn by using the touch table. By being part of the research you can help us improve the exhibit.
**What happens in the research?**

- You (and your child or children) will use the touch table as you normally would. You can stay at the table for as long or as little as you want.
- Researchers will take notes as they watch and listen to what you do and say while at the table. This helps us understand how people use the exhibit.
- When you are done, you (or your child or children) might be asked to fill out a survey about what you learned and what you thought of the exhibit. If you prefer, we can ask you the questions and write down your answers for you.

**What we do with the information?** Only project staff will see the raw data. We will not record any identifying information about you. We will write about what we learn for educators, exhibit developers, and our funders. We may publish in journals or online.
**Are there any benefits?** There are no direct benefits to you, though you may feel good about helping us improve the exhibit.
**Are there any risks?** You might be embarrassed by something you say or do. Because we don't record your identity, there are no further risks.
**Do I have to take part?** No. However, if you choose to use the touch table *at this time*, an observer will take notes about your actions and words. If you don't want to be part of the research, please use the other table, or come back later when the research is not going on.
**What if I have more questions?** The researcher at the exhibit may be able to answer your questions, or you can talk to staff in the museum's main office. Or call the lead evaluator, Jim Hammerman, at 617-873-9600, jim_hammerman@terc.edu. You can also call Polly Hubbard on TERC's oversight board for human subjects research (IRB) at 617-873-9600 or polly_hubbard@terc.edu.
YES. I understand the *Life on Earth* evaluation and agree for me and/or my child/ children to be part of the study and survey.

______________________________ ____________________
Signature Date

**Thank you for helping us improve this exhibit!**

## Demographic Information

The following tables provide detailed demographic information about study participants.

### *DeepTree Video Participants*

Table 23: DeepTree Video Observation Sample by Age Category and Gender

| Age Category | Male | Female | Total |
|---|---|---|---|
| Young child | 0 | 5 | 5 (3%) |
| Child | 34 | 13 | 46 (27%) |
| Teen | 10 | 5 | 15 (9%) |
| Age 20-39 | 34 | 41 | 75 (44%) |
| Age 40-59 | 12 | 8 | 20 (12%) |
| 60+ | 5 | 2 | 7 (4%) |
| **Total** | **95 (56%)** | **74 (44%)** | **169** |

Note: N=170. One person's age/gender category was not ascertained from video observation coding.

Table 24: DeepTree Video Study Survey Sample by Self-reported Age Category and Gender

| Age Category | Male | Female | (No gender reported) | Total |
|---|---|---|---|---|
| (no age reported) | 0 | 0 | 2 | 2 (2%) |
| Younger than 9 | 4 | 0 | 1 | 5 (4%) |
| 9-14 years | 17 | 12 | 2 | 31 (23%) |
| 15-23 years | 10 | 5 | 1 | 16 (12%) |
| 24-39 years | 12 | 11 | 0 | 23 (17%) |
| 40-59 years | 19 | 20 | 1 | 40 (30%) |
| 60+ years | 7 | 5 | 3 | 15 (11%) |
| **Total** | **69 (52%)** | **53 (40%)** | **10 (8%)** | **132** |

Note: N=132.

Visitors' primary language: About 84% of participants in the DeepTree Video study (n = 111) indicated that they spoke English at home, with over 25% of those (n = 29) speaking another language as well, including Spanish (n=13, 9.8% of total sample), Chinese (n = 8; 6.0%), French (n = 5, 3.8%), Tagalog (n = 2, 1.5%), as well as Russian, Arabic, Italian and Latin (n = 1 for each, 0.8%). Of the 16% (n = 21) who did not indicate that they spoke English at home, languages included French (n = 7, 5.3%), Japanese (n = 3, 2.2%), Chinese (n = 3, 2.2%), Italian (n = 2, 1.5%), and German (n = 2, 1.5%). Four (n = 4, 3%) participants did not indicate any language(s) they spoke at home.

Table 25: Self-identified Race/Ethnicity of DeepTree Video Observation Survey Sample

| | n (Percentage) |
|---|---|
| **Race*** | |
| White | 96 (73%) |
| Asian American | 16 (12%) |
| Black | 2 (2%) |
| Pacific Islander | 5 (4%) |
| Native American | 1 (1%) |
| Other* | 9 (7%) |
| Did not report | 8 (6%) |
| **Ethnicity** | |
| Hispanic or Latino/a | 9 (7%) |

Notes: N=132. Five (4%) individuals chose multiple racial categories. When identifying another ethnicity, participants reported Asian, Italian, Japanese, and Middle-Eastern.

Table 26: DeepTree Video Participant-reported Education by Age Category

| | Some Elem. School | Some Middle school | Some High school | High school diploma | Some college | 2yr college degree | 4yr college degree | Master's degree | Prof. degree | Other |
|---|---|---|---|---|---|---|---|---|---|---|
| < 9 yrs | 4 | | | | | | | | | |
| 9-14 yrs | 11 | 14 | 1 | | | | | | | 4 |
| 15-23 yrs | | | 1 | 2 | 6 | 0 | 3 | | | 4 |
| 24-39 yrs | | | | | 2 | 1 | 10 | 5 | 1 | 4 |
| 40-59 yrs | | | | 1 | 1 | 1 | 23 | 8 | 3 | 3 |
| 60+ yrs | | | | | | | 6 | 6 | 2 | 1 |
| **Total** | **15** | **14** | **2** | **3** | **9** | **2** | **42** | **19** | **6** | **16** |

Note: N=132.

### *DeepTree Naturalistic Participants*

Table 27: DeepTree Naturalistic Observation Sample by Age Category and Gender

| Age Category | Male | Female | Uncertain | Total |
|---|---|---|---|---|
| Young child | 17 | 4 | 3 | 24 (7%) |
| Child | 42 | 35 | 0 | 77 (24%) |
| Teen | 24 | 19 | 0 | 43 (13%) |
| Age 20-39 | 64 | 67 | 0 | 131 (40%) |
| Age 40-59 | 14 | 15 | 0 | 29 (9%) |
| Age 60+ | 5 | 7 | 0 | 12 (4%) |
| Uncertain Age & Gender | | | | 10 (3%) |
| **Total (Pct of known)** | **166 (52.5%)** | **147 (46.5%)** | **3 (0.9%)** | **326** |

Note: N=326.

Table 28: DeepTree Naturalistic Study Survey Sample by Self-reported Age Category and Gender

| Age Category | Male | Female | (No gender reported) | Total |
|---|---|---|---|---|
| Younger than 9 | 0 | 0 | 0 | 0 |
| 9-14 years | 4 | 1 | 0 | 5 (15%) |
| 15-23 years | 2 | 7 | 0 | 9 (27%) |
| 24-39 years | 3 | 3 | 0 | 6 (18%) |
| 40-59 years | 4 | 6 | 1 | 11 (33%) |
| 60+ years | 2 | 0 | 0 | 2 (6%) |
| **Total** | **15 (45%)** | **17 (52%)** | **1 (3%)** | **33** |

Note: N=33.

Table 29: Self-identified Race/Ethnicity of DeepTree Naturalistic Survey Sample

| | n (Percentage) |
|---|---|
| **Race** | |
| White | 21 (64%) |
| Asian American | 5 (15%) |
| Black | 0 |
| Pacific Islander | 0 |
| Native American | 2 (6%) |
| Other* | 6 (18%) |
| **Ethnicity (of n = 29 reporting)** | |
| Hispanic or Latino/a | 4 (12%) |

Note: N=33. One participant indicated multiple categories; when identifying another ethnicity, participants reported Latin, Jewish, and Middle-Eastern.

Table 30: DeepTree Naturalistic Participant-reported Education by Age Category

| | Some Elem. School | Some Middle school | Some High school | High school diploma | Some college | 4yr college degree | Master's degree | Prof. degree | Other |
|---|---|---|---|---|---|---|---|---|---|
| 9-14 yrs | 1 | 3 | 1 | | | | | | |
| 15 – 23 yrs | | | | 1 | 5 | 3 | | | |
| 24-39 yrs | | | | | 2 | 1 | 3 | | |
| 40 - 59 yrs | | | | | 1 | 4 | 1 | 2 | 3 |
| 60+ yrs | | | | | 1 | | 1 | | |
| **Total** | **1** | **3** | **1** | **1** | **9** | **8** | **5** | **2** | **3** |

Note: N=33.

## *BAT Video Participants*

Table 31: BAT Video Study Sample by Age Category and Gender

| Age Category | Male | Female | Uncertain | Total |
|---|---|---|---|---|
| Young child | 0 | 2 | 0 | 2 (9%) |
| Child | 1 | 1 | 1 | 3 (13%) |
| Teen | 0 | 0 | 0 | 0 |
| Age 20-39 | 5 | 5 | 1 | 11 (48%) |
| Age 40-59 | 3 | 1 | 0 | 4 (17%) |
| 60+ | 0 | 0 | 0 | 0 |
| Uncertain Age | 3 | 0 | 0 | 3 (13%) |
| **Total** | **12 (52%)** | **9 (39%)** | **2 (9%)** | **23** |

Note: N=23.

Table 32: BAT Video Study Survey Sample by Self-reported Age Category and Gender

| Age Category | Male | Female | Total |
|---|---|---|---|
| Younger than 9 | 0 | 1 | 1 (5%) |
| 9-14 years | 2 | 0 | 2 (11%) |
| 15-23 years | 4 | 3 | 7 (37%) |
| 24-39 years | 4 | 1 | 5 (26%) |
| 40-59 years | 3 | 1 | 4 (21%) |
| 60+ years | 0 | 0 | 0 |
| **Total** | **13 (68.4%)** | **6 (31.6%)** | **19** |

Note: N=19.

Table 33: BAT Video study Self-identified Race and Ethnicity of Survey Sample

| | n (Percentage) |
|---|---|
| **Race** | |
| White | 8 (42.1%) |
| Asian American | 7 (36.8%) |
| Black | 0 |
| Pacific Islander | 0 |
| Native American | 2 (10.5%) |
| Other* | 2 (10.5%) |
| Did not report | 2 (10.5%) |
| **Ethnicity** | |
| Hispanic or Latino/a | 8 (42%) |

Note: N=19. Two participants indicated multiple categories; when identifying another ethnicity, participants reported El Salvadorian and Middle-Eastern.

Table 34: BAT Video Study Participant-reported education by Age Category

| | Some Elem. School | High school diploma | Some college | 2yr college degree | 4yr college degree | Master's degree | Prof. degree | Other |
|---|---|---|---|---|---|---|---|---|
| < 9 yrs | 0 | | | | | | | 1 |
| 9-14 yrs | 2 | | | | | | | 0 |
| 15 – 23 yrs | | 1 | 3 | 0 | 2 | | | 1 |
| 24-39 yrs | | | | | 1 | 3 | | 0 |
| 40-59 yrs | | | | 2 | 1 | | | 1 |
| 60+ yrs | | | | | | | | 0 |
| **Totals** | **2** | **1** | **3** | **2** | **4** | **3** | **0** | **3** |

Note: N=19. One person did not report education information.

### *BAT Naturalistic Participants*

Table 35: BAT Naturalistic Study Sample by Age Category and Gender

| Age Category | Male | Female | Total |
|---|---|---|---|
| Young child | 14 | 3 | 17 (11%) |
| Child | 15 | 17 | 32 (21%) |
| Teen | 6 | 9 | 15 (10%) |
| Age 20-39 | 31 | 34 | 65 (42%) |
| Age 40-59 | 5 | 5 | 10 (6%) |
| 60+ | 3 | 2 | 5 (3%) |
| Uncertain Age & Gender | | | 12 (8%) |
| **Total** | **74 (51.4%)** | **70 (48.6%)** | **156** |

Note: N=156

Table 36: BAT Naturalistic Study Survey Sample by Self-reported Age Category and Gender

| Age Category | Male | Female | Total |
|---|---|---|---|
| Younger than 9 | 1 | 1 | 2(17%) |
| 9-14 years | 1 | 0 | 1 (8%) |
| 15-23 years | 0 | 2 | 2(17%) |
| 24-39 years | 0 | 2 | 2(17%) |
| 40-59 years | 1 | 4 | 5(42%) |
| 60+ years | 0 | 0 | 0 |
| **Total** | **3 (25%)** | **9 (75%)** | **12** |

Note: N=12.

Table 37: BAT Naturalistic Study Self-identified Race and Ethnicity of Survey Sample

| | n (Percentage) |
|---|---|
| **Race** | |
| White | 9 (75%) |
| Asian American | 1 (8%) |
| Black | 0 |
| Pacific Islander | 0 |
| Native American | 0 |
| Other* | 2 (17%) |
| **Ethnicity (of 11 reporting)** | |
| Hispanic or Latino/a | 1 (8%) |

Note: N=12. When identifying another ethnicity, one participant reported Mayan.

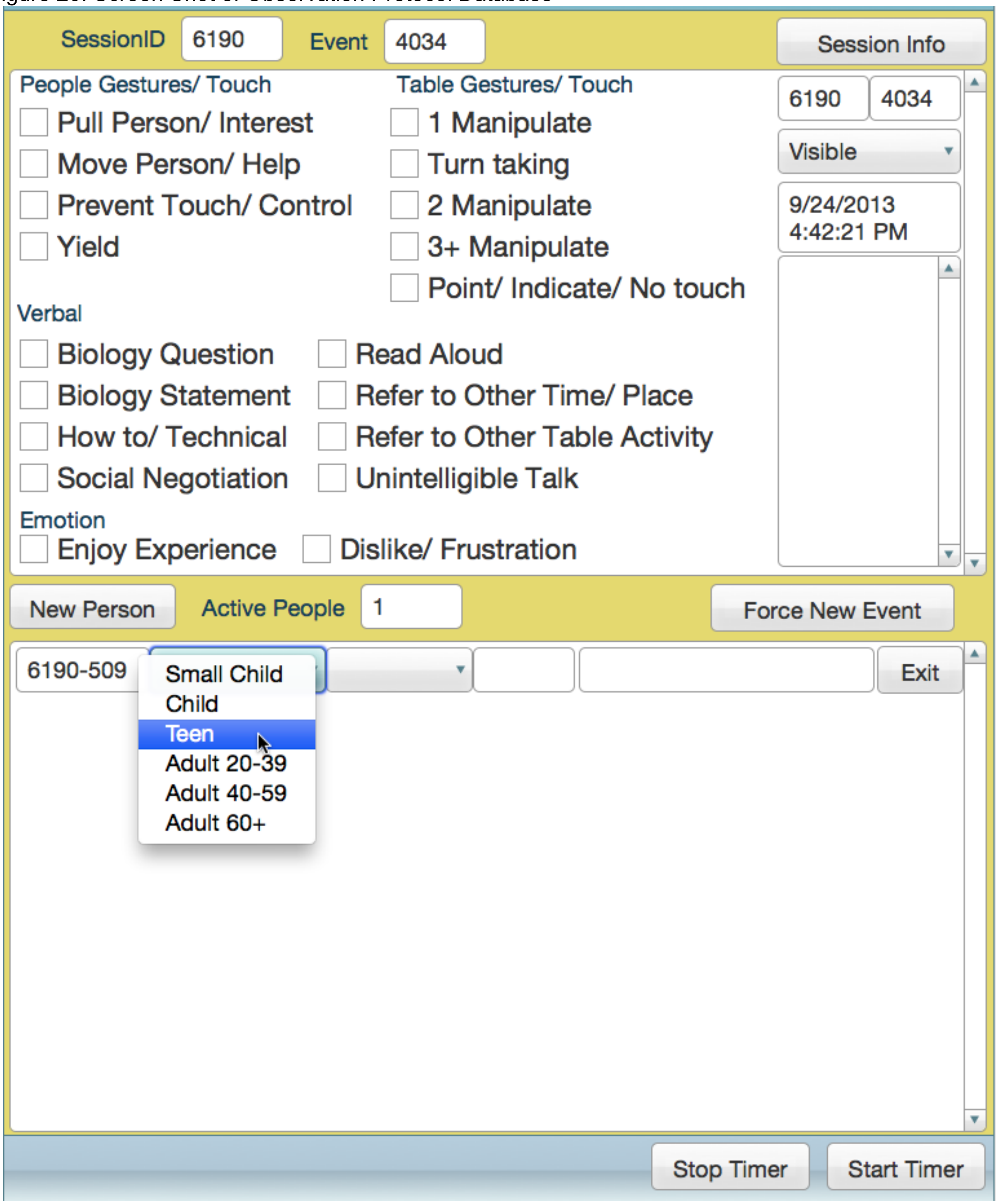
Table 38: BAT Naturalistic Study Participant-reported education by Age Category

| | Some Elem. School | High school diploma | Some college | 2yr college degree | 4yr college degree | Master's degree | Prof. degree | Other |
|---|---|---|---|---|---|---|---|---|
| < 9 yrs | 2 | | | | | | | 0 |
| 9-14 yrs | 1 | | | | | | | 0 |
| 15-23 yrs | | | | | 2 | | | 0 |
| 24-39 yrs | | | | | 1 | 1 | | 0 |
| 40-59 yrs | 1 | 0 | 0 | 0 | 0 | 1 | 2 | 1 |
| 60+ yrs | | | | | | | | 0 |
| **Totals** | **4** | **0** | **0** | **0** | **3** | **2** | **2** | **1** |

Note: N=12.

## Observation Protocol Software

Figure 20: Screen Shot of Observation Protocol Database




Note: The open field in the top half was for brief observational comments.
In the bottom half, the first drop-down menu displayed Age categories (as shown); the second drop-down menu provided Gender options (Male, Female, Uncertain). The third field was for Sticker ID. The fourth for descriptive comments.

## Observation Protocol Codebook

# Life on Earth

### Touch-Table Behavior Observation Categories

Unless otherwise noted, the below sets of behaviors will be coded at the group level, across a set period of time, meaning that each time any of the actions occurs within the group during the time period, it will be coded. Thus, each behavior will be marked the first time it occurs during every [20 second] coding time period, rather than marked every instance. Group level coding will ensure that important elements of the group behavior around the table will be captured. Try to code when an event occurs but if you're about to code and the record updates, it's better to code in the next event than not at all.

The session protocol background on the iPad will turn yellow during active coding (with 20-second event periods automatically being generated). When not active, background is blue. Protocol becomes active when new person is added to empty table, when "Start Timer" is selected, or when the record shows that there is a person at the table and "Start Observing" is pressed from the Session screen.

## Nonverbal

**People Gestures/ Touch**

1. Pull Person /Interest: A person pulls the body of another person over to the table, is pulled by another person, or gestures to another person to encourage him/her to come to the table (may also include verbal behavior, which should be coded as well, for example, "Social Negotiation" or "Enjoy Experience" or both). The focus of this code is physical expressions of engagement and interest.
2. Move Person /Help: A person moves another person or part of their body in such a way to assist that person's manipulation of the table; or the subject's body is moved in such a way by another person. Both people are at the exhibit. (e.g., adult lifts a child, moves their hand to specific places). The focus of this code is helping/aiding use.
3. Prevent Touch/ Control: This code combines aggression/hitting/pushing with actions to constrain or restrain another from touching the table. One person either touches another person forcibly, by hitting, shoving, pushing, or blocking; or moves another person so that they cannot touch the table by carrying or lifting him/her away, pulling him/her away or holding his/her hands or body so s/he cannot touch the table. The focus of this code is physical action that expresses social control of behavior.
4. Yield: A person yields control or operation of the table to another person, by removing their own hands from the table, nodding, gesturing, or otherwise indicating that the other(s) should go ahead. The focus of this code is physical action expressing positive/ collaborative social negotiation. This "Yield" code has been difficult to reliably observe, since it can be subtle. With the inclusion of the code "Turn Taking," this "Yield" code may be less important.

**Table Gestures/ Touch**

5. 1 Manipulate: One person touches table or moves icon(s) around or opens item(s) or otherwise elicits change in the table. Pointing in the vicinity of the table (within 3 inches) that is associated with a touch (as far as we can tell) counts as manipulation as well.

6. Turn Taking: Two or more people alternate touching table and/or moving icons and/or opening items and/or otherwise eliciting change in the table. Note: if one person touches the table, and then a second person touches the table, both 1 manipulate and Turn Taking are coded, but if this type of alternating is going on, "Turn taking" *can* be the first code in a new time interval.
7. 2 Manipulate: Two people simultaneously touch the table and/or move icons around and/or open items and/or otherwise elicit change in the table. Pointing in the vicinity of the table (within 3 inches) that is associated with a touch (as far as we can tell) counts as manipulation as well.
8. 3+ Manipulate: Three or more people simultaneously touch the table and/or move icons around and/or open items and/or otherwise elicit change in the table. Pointing in the vicinity of the table (within 3 inches) that is associated with a touch (as far as we can tell) counts as manipulation as well.
9. Point/ Indicate/ No touch: Person points finger at something on the table, indicating attention, but doesn't touch and manipulate table. Verbal comments may help clarify the intention/ purpose of the pointing.

## Verbal

**Verbal**

10. Biology Question: Person asks a question about science content including (but not limited to) information about biology or evolution (natural selection, common descent, tree of life, heredity, change over time, population changes, preferential survival, etc.) or causal mechanisms for these such as geological change that caused speciation.
11. Biology Statement: Person answers a question and/or makes a comment about science content including (but not limited to) information about biology or evolution (natural selection, common descent, tree of life, heredity, change over time, population changes, preferential survival, etc.) or causal mechanisms for these such as geological change that caused speciation. This may include reading aloud science content from the table (both Biology Statement and Read Aloud would be coded in that instance).
12. How to/ Technical: Person asks for help or provides an explanation of how to manipulate the table, or about what to do at the table, or about how the table works (e.g. "How do you do this?" or "What am I supposed to do here?" or "How do you work this?" or "Try pressing that button"). This code emphasizes discussion about *how* to do something. Statement may be coded both here and as Social Negotiation if both are present ("Can you do that again?"). If people are reading instructions about how to do something, then both How to/ Technical and Read Aloud would be coded.
13. Social Negotiation: Person asks or talks with another person about social interactions around manipulating the table (e.g. "Let's both choose an animal "or "Let's see if we can stop it [the FloTree]") or space around the table (e.g. "move over!"). Person tells another person to manipulate the table by one or more specific actions (e.g. "Pull that picture over here") or to continue doing what s/he is doing at the table, or that it is alright to do what s/he is doing (e.g. "yes" or "You've got it") or to stop doing what s/he is doing (e.g., "Stop it!" or "No!") or to look at a particular image or text. This code emphasizes the negotiation about *who should do what*. Statement may be coded both here and as How to/Technical if both are present ("Can you do that again?"). Comments made to negotiate leaving the table (e.g. "Ready to go?") should be coded here.

14. Read Aloud: Person looks at the table graphics or text and verbally repeats them to self or another person. Even a single word read aloud (e.g. "DNA" or "Restart") should be coded here. Since we can't determine whether someone silently reading is actually reading or just looking, we require verbalization for this code. Observer must clearly hear spoken words and not just surmise that reading aloud is occurring. Note: If content read aloud is Biology Statement or How to/ Technical, these should also be coded.
15. Refer to Other Time/ Place: Person makes a comment about a past or future experience with another museum exhibit (e.g., "Can we find that big fish we saw over there?") or makes a comment about the exhibit that relates it to an object or experience in the subject's home, school, or community (e.g., "This looks like the big tree in the playground,"), or relates to other phenomena, objects, or experiences outside of the museum (e.g. "We learned about this at school," or "We saw one of these at the zoo!").
16. Refer to Other Table Activity: Subject makes a comment that describes a past or future interaction with the table.
17. Unintelligible Talk: Subject speaks in a foreign language or in an inaudible manner such that talking is occurring but observer is unable to discern any meaning.

**Emotion**

18. Enjoy Experience: Person describes or expresses pleasure or enjoyment in their experience at the table (e.g. Laughter, "I love seeing the tree zoom by" or "I want to see another one" or "Doing this together is fun" or "This is so cool!"), verbally expresses engagement with content (e.g. "wow!"), or tells another person to come to the exhibit (e.g. "You have to come look at this!"). Just smiling is not enough to be coded here; there should be other overt expressions of enjoyment.
19. Dislike/ Frustration: Person describes dislike or annoyance or frustration toward the table or table experience (e.g. "I don't like this," "I don't want to do this anymore," "I'm bored," or "This is too hard," or, in particular, "this zooming makes me dizzy" or other language or behaviors indicating that dizziness).

## Open Notes

A place to (BRIEFLY) note other relevant behaviors, utterances (include direct quotes where possible) or patterns that seem interesting to the observer and are not captured in the regular codes. Notes are intended primarily as reminders to be fleshed out in qualitative comments between coding sessions. Because of the nature of the coding, these notes may be recorded in later event periods than those in which they occur, or may be partially recorded in one event, then continued in the next. Try not to miss codes while taking notes, and we'll expect that your coding isn't as accurate when there's a note in this field.

If noting a quote, start the quote with a "**q**" to indicate that what follows is a direct quote. (e.g., q This is so cool; or q so cool)

If noting that some people at the table are watching and apparently waiting for their turn at the table, use a "**w**" or "**wait**" to indicate this is occurring. If/ when they decide to interact with the table, use "**j**" or "**join**" to indicate a change in behavior.

If events become too complex and/or frequent to code, indicate that coding is incomplete by putting "**inc**" in the open notes. When it's settled down again, use "**OK**" in the notes.

At the end of a coding session, you can review these notes by using the "Session Events" button, and then more complete field notes about the session should be noted in the Session Notes section.

## Person Arrival/ Departure Information

When a person arrives at the table, pressing the "**New Person**" button will record their arrival time. Doing this is essential to start the clock on their experience at the table, even if you wait a bit to fill in demographics. Basic demographic information will be recorded as follows:

20. Age Category: Select best estimate from the following:
    *Young child (< ~6 yrs); Child; Teen; Adult 20-39; Adult 40-59; Adult 60+*
21. Gender: Select best estimate from the following: *Male; Female; Uncertain*
22. Subject Number: If subject has a sticker number (videotaped only) note this in the number space.
23. Open Description: Record other observable personal information —characteristics of dress or body that will enable you to link the person to this identifying record. These will not be coded in analysis but are just used to specify this person and make them easier to identify when coding departure.

Pressing the "**Exit"** button will record the person's time of departure from the table. This is also crucial to do in a timely fashion so that we can analyze dwell time at the exhibit.

**Mistaken person**: If you mistakenly add a person, type "**oops**" in the Open Description space before pressing Exit so that we'll discount that from the data corpus.

**Returning Person**: We don't expect observers to notice if a person returns to the table since that could happen over a varied time period. However if you do notice someone who seems to have returned, that can be coded in the Subject Number or with an "**r**" or the phrase "**ret**" in the Open Description category for their second appearance at the table. If they don't have a sticker (non-video) and you can add their previous ID number to this Open Description, but we're not making it easy to find that so don't sweat it if you can't.

## Observe People/Observe Events/Observe Both

The session protocol can be configured to observe in three different modes. If two observers are present, they indicate their coding partner by noting the other observer's session number in the place indicated, with one indicating s/he will "observe events" and the other indicating s/he will "observe people." They will each be provided with the appropriate corresponding screen to code people arriving and departing or events occurring, and the data will be linked. If only one observer is present, this person signs into the session and indicates that s/he will "observe both" (visitors arriving and departing exhibit *and* events occurring at the table).

## Observation Protocol Inter-Rater Reliability

The complex coding scheme and process meant ascertaining inter-rater reliability was somewhat difficult. While average percent agreement among the five evaluator raters (Jim, Amy, Jon, Anita, and Julie) was 90%, this statistic masks the effects both of chance agreement (more prominent when the prevalence of coding for an event is either very high or very low) and of differences in reliability by specific code. The Cohen's Kappa statistic is designed to account for chance agreement – it calculates the proportion of times raters agree beyond what would be expected by chance. Overall, this statistic was considered "fair" or "moderate" ($\kappa = .40$). Cohen's Kappa is known to be low when overall prevalence of a code is high or low; calculating Kappa just for the codes whose prevalence was between 10% and 90% yields a slightly higher but still marginal $\kappa = .45$.

As we discussed these moderate but not robust reliability statistics, we realized that some of our procedures might be affecting the extent to which raters were recorded as agreeing, even if they were seeing the same thing. In particular, our protocol for coding in 20 second intervals was that if an event occurred towards the end of an interval and the observer hadn't had time to code it before the software began a new coding interval, the observer was to code the event in the next interval. Thus, differences in speed and order in which observers took note of various observed events might lead the same event to be coded in different time intervals. Slight differences in when the 20 second intervals began could also put codes in adjacent intervals, substantially reducing Kappa, even if raters were observing and coding the same things.

We tried accounting for this problem a few different ways. First, we essentially doubled the length of the interval, to 40 seconds, and calculated Kappa in these larger intervals. We didn't know which rater might be "ahead" of the other, so we combined intervals in two ways—either starting at odd numbered intervals (merging intervals 1 and 2, then 3 and 4, etc.) or starting at even numbered intervals (merging intervals 2 and 3, then 4 and 5, etc.). Events that crossed the boundary of these 40 second intervals would still be considered disagreements, but by calculating both ways, we could see if the values differed markedly.

We also calculated a more complicated "Next Kappa" statistic that kept 20 second intervals, but allowed for agreement if coders agreed either in the current interval or, when they disagreed in the current interval, then if they agreed across the coding interval boundary. Cross-boundary agreement was calculated first with Coder A lagging, then with Coder B lagging; and the maximum of these two was reported. This new statistic gave slightly higher values than either of the doubled interval coded values (average for moderately prevalent codes, all raters: Next Kappa = .63; Double Kappa1 = .50; Double Kappa2 = .56) which was encouraging since, though it was more complicated to compute, we felt Next Kappa better reflected the problem we were trying to resolve than the double interval solution.

In addition, lead evaluator Jim Hammerman from TERC, senior evaluator Amy Spiegel from UNL, and field evaluation team leader Anita Smith tended to be more reliable with one another than were other researchers on the team (for moderately prevalent codes, average Next Kappa: Jim-Amy: .68; Jim-Amy-Anita: .67). In addition, Next Kappa values for individual event codes for Jim-Amy-Anita were all above .60 except for Turn-Taking (.57), Bio Question (.59) and Enjoyment (.51). Original Cohen's Kappa values for individual event codes for Jim-Amy-Anita were all above .40 except for 1 Manipulate (.25) and Turn-taking (.36), for which Jim and Amy had high levels of agreement, but which could be confused if an observer tracked alternating table touches differently. For codes which had high or low prevalence, Byrt, Bishop & Carlin's Kappa correction (Byrt et al., 1993) cited in (Hallgren, 2012) gives values $\geq .94$ for all raters and in particular for Jim, Amy and Anita. Though not ideal, we felt that with original levels of agreement that were “fair to moderate” and levels of agreement with our modified statistic that could be considered “substantial,” the team of Amy, Jim and Anita were sufficiently reliable scorers of our event-coding instrument to proceed with data collection for the naturalistic study, where there would be no video back-up.

When possible, we also checked inter-rater reliability after the fact. For the DeepTree Video Study, Amy coded all the videotaped events and Jim coded 20% of them. Although video alignment wasn’t perfect which could contribute to error variation and somewhat reduced reliability statistics (though in over 90% of intervals, misalignment was 3 seconds or fewer), re-checking reliability in these double-coded events we find Next Kappa values for moderately prevalent codes averaging .62, with individual values all above .60 except for Point (.40), Social Negotiation (.49) and Enjoyment (.57). The Byrt et al. prevalence-adjusted Kappa for high and low prevalence codes are all above .91 except for Other Time/ Place (.84). While slightly lower than during the time we were initially establishing reliability, these values are sufficiently high to use, though we will be cautious about the Point and Negotiation codes.

## Dwell Time Details

### *By Age Group*

The following tables provide a range of statistics about dwell time by age group.

Table 39: Time Spent at Table by Age Category for DeepTree Video study

| Age Category | Count | Mean (from log transformed data) | Median | Middle 50% |
|---|---|---|---|---|
| Young child | 5 | 296 s (4.93 min) | 368 s (6.13 min) | 3.28 – 7.60 min |
| Child | 46 | 446 s (7.44 min) | 509 s (8.48 min) | 4.85 – 12.18 min |
| Teen | 15 | 510 s (8.51 min) | 580 s (9.67 min) | 6.65 – 10.92 min |
| Age 20-39 | 75 | 368 s (6.13 min) | 351 s (5.85 min) | 3.75 – 10.63 min |
| Age 40-59 | 20 | 440 s (7.33 min) | 742 s (12.37 min) | 5.85 – 13.83 min |
| Age 60+ | 7 | 262 s (4.37 min) | 287 s (4.78 min) | 1.60 – 13.63 min |
| **Total** | **169** | **400 s (6.67 min)** | **455 s (7.58 min)** | **4.32 – 11.92 min** |

Table 40: Time Spent at Table by Age Category for DeepTree Naturalistic study

| Age Category | Count | Mean (from log transformed data) | Median | Middle 50% |
|---|---|---|---|---|
| Young child | 24 | 44 s (0.73 min) | 38 s (0.64 min) | 22 s – 1.58 min |
| Child | 77 | 55 s (0.92 min) | 46 s (0.77 min) | 20 s – 2.87 min |
| Teen | 43 | 63 s (1.05 min) | 54 s (0.90 min) | 21 s – 3.42 min |
| Age 20-39 | 131 | 90 s (1.50 min) | 104 s (1.73 min) | 46 s – 3.53 min |
| Age 40-59 | 29 | 105 s (1.75 min) | 112 s (1.87 min) | 44 s – 3.68 min |
| Age 60+ | 12 | 70 s (1.17 min) | 61 s (1.02 min) | 26 s – 3.62 min |
| **Total** | **326** | **69 s (1.14 min)** | **67 s (1.12 min)** | **26 s – 3.23 min** |

Table 41: Time Spent at Table by Age Category for BAT Video study

| Age Category | Count | Mean (from log transformed data) | Median | Middle 50% |
|---|---|---|---|---|
| Young child | 2 | 527 s (8.78 min) | 527 s (8.78 min) | – |
| Child | 3 | 1215 s (20.25 min) | 1141 s (19.02 min) | – |
| Teen | 0 | – | – | – |
| Age 20-39 | 11 | 385 s (6.41 min) | 364 s (6.07 min) | 5.30 – 7.87 min |
| Age 40-59 | 4 | 616 s (10.27 min) | 718 s (11.97 min) | 3.89 – 25.65 min |
| 60+ | 0 | – | – | – |
| Unknown | 3 | 403 s (6.72 min) | 455 s (7.58 min) | – |
| **Total** | **23** | **502 s (8.36 min)** | **424 s (7.07 min)** | **6.05 –11.32 min** |

Table 42: Time Spent at Table by Age Category for BAT Naturalistic study

| Age Category | Count | Mean (from log transformed data) | Median | Middle 50% |
|---|---|---|---|---|
| Young child | 17 | 81 s (1.35 min) | 60 s (1.00 min) | 27 s – 4.16 min |
| Child | 32 | 143 s (2.38 min) | 134 s (2.23 min) | 1.06 – 7.09 min |
| Teen | 15 | 118 s (1.97 min) | 106 s (1.77 min) | 39 s – 5.87 min |
| Age 20-39 | 66 | 112 s (1.86 min) | 143 s (2.39 min) | 56 s – 4.08 min |
| Age 40-59 | 9 | 138 s (2.30 min) | 91 s (1.52 min) | 1.03 – 7.79 min |
| 60+ | 5 | 44 s (0.73 min) | 56 s (0.93 min) | 26 s – 1.49 min |
| Unknown | 12 | 35 s (0.58 min) | 18 s (0.30 min) | 10 s – 4.29 min |
| **Total** | **156** | **102 s (1.71 min)** | **105 s (1.76 min)** | **44 s – 4.35 min** |

### *By Group Size*

The following tables provide a range of statistics about dwell time by group size.

Table 43: Modal Number of People at Exhibit vs. Dwell Time for DeepTree Video Study

| Group Size | Count | Mean (from log transformed data) | Median | Middle 50% |
|---|---|---|---|---|
| 1 person | 10 | 295 s (4.91 min) | 322 s (5.36 min) | 3.38 – 6.37 min |
| 2 people | 46 | 481 s (8.02 min) | 540 s (9.00 min) | 3.53 – 16.23 min |
| 3 people | 52 | 401 s (6.68 min) | 465 s (7.76 min) | 4.68 – 10.58 min |
| 4 people | 39 | 336 s (5.60 min) | 351 s (5.85 min) | 3.75 – 9.47 min |
| 5 people | 15 | 826 s (13.77 min) | 955 s (15.92 min) | 12.18 – 17.18 min |
| 6 people | 2 | 110 s (1.84 min) | 110 s (1.84 min) | |
| 7 people | 3 | 25 s (0.42 min) | 16 s (0.27 min) | |
| **Total** | **169** | **400 s (6.67 min)** | **455 s (7.58 min)** | **4.32 – 11.92 min** |

Note: There were too few groups with more than 5 people to calculate some statistics.

Table 44: Number of People at Exhibit vs. Dwell Time for DeepTree Naturalistic study

| Group Size | Count | Mean (from log transformed data) | Median | Middle 50% |
|---|---|---|---|---|
| 1 person | 48 | 68 s (1.13 min) | 52 s (0.87 min) | 25 s – 3.60 min |
| 2 people | 133 | 85 s (1.42 min) | 82 s (1.37 min) | 34 s – 3.69 min |
| 3 people | 83 | 55 s (0.91 min) | 48 s (0.80 min) | 25 s – 2.08 min |
| 4 people | 24 | 85 s (1.41 min) | 107 s (1.79 min) | 33 s – 3.80 min |
| 5 people | 12 | 64 s (1.06 min) | 147 s (2.44 min) | 9 s – 4.82 min |
| 6 people | 8 | 103 s (1.71 min) | 102 s (1.70 min) | 1.50 – 1.91 min |
| 7 people | 5 | 195 s (3.24 min) | 171 s (2.85 min) | 2.62 – 4.60 min |
| **Total** | **326** | **69 s (1.14 min)** | **67 s (1.12 min)** | **26 s – 3.25 min** |

Note: There are 13 people in the Naturalistic study for whom we do not have group size information. Median time at the table for these people was 12 s.

Table 45: Number of People at Exhibit vs. Dwell Time for BAT Video Study

| Group Size | Count | Mean (from log transformed data) | Median | Middle 50% |
|---|---|---|---|---|
| 1 person | 5 | 447 s (7.46 min) | 471 s (7.87 min) | 4.33 – 13.83 min |
| 2 people | 12 | 646 s (10.76 min) | 546 s (9.09 min) | 7.05 – 19.22 min |
| 3 people | 2 | 264 s (4.39 min) | 264 s (4.39 min) | – |
| 4 people | 4 | 374 s (6.23 min) | 364 s (6.07 min) | 6.05 – 6.59 min |
| **Total** | **23** | **502 s (8.36 min)** | **424 s (7.07 min)** | **6.05 – 11.32 min** |

Table 46: Number of People at Exhibit vs. Dwell Time for BAT Naturalistic Study

| Group Size | Count | Mean (from log transformed data) | Median | Middle 50% |
|---|---|---|---|---|
| 1 person | 24 | 72 s (1.20 min) | 62 s (1.04 min) | 18 s – 4.97 min |
| 2 people | 50 | 110 s (1.83 min) | 134 s (2.23 min) | 53 s – 4.28 min |
| 3 people | 47 | 109 s (1.81 min) | 100 s (1.67 min) | 43 s – 5.13 min |
| 4 people | 18 | 100 s (1.67 min) | 90 s (1.51 min) | 55 s – 4.26 min |
| 5 people | 6 | 171 s (2.84 min) | 264 s (4.40 min) | 2.63 – 6.76 min |
| 6 people | 10 | 126 s (2.10 min) | 115 s (1.91 min) | 1.48 – 3.50 min |
| **Total** | **156** | **102 s (1.71 min)** | **105 s (1.76 min)** | **44 s – 4.35 min** |

Note: There is 1 person in the people in the BAT Naturalistic study for whom we do not have group size information.

## Event Proportions by Intervals

The following tables provide information about proportion of coded events which showed various engagement behaviors, as a function of group size.

### *DeepTree Event Proportions*

Table 47: Proportion of Engagement Behaviors (SD) over Total Available Time Intervals by Group Size for DeepTree Video Study

| Behavior | 1 person (n = 10) | 2 people (n = 46) | 3 people (n = 52) | 4 people (n = 39) | 5+ people * (n=20) |
|---|---|---|---|---|---|
| Pull Person/ Interest | | | 0.3% (1%) | | |
| Move Person/ Help | | 1% (6%) | 0.2% (1%) | | 0.2% (0.5%) |
| Prevent Touch/ Control | 0.4% (1%) | 4% (10%) | 2% (4%) | 13% (22%) | 6% (6%) |
| Yield | | 0.2% (0.6%) | 2% (5%) | | 0.5% (0.8%) |
| 1 Manipulate | 91% (5%) | 92% (5%) | 86% (15%) | 89% (9%) | 92% (11%) |
| Turn Taking | 6% (17%) | 21% (12%) | 16% (11%) | 21% (15%) | 26% (16%) |
| 2 Manipulate | 4% (9%) | 24% (20%) | 37% (26%) | 36% (22%) | 46% (11%) |
| 3+ Manipulate | | 0.3% (2%) | 5% (10%) | 12% (13%) | 15% (17%) |
| Point/ Indicate | 2% (2%) | 23% (15%) | 26% (17%) | 23% (16%) | 28% (17%) |
| Biology Question | | 6% (1%) | 8% (1%) | 5% (1%) | 0.7% (1%) |
| Biology Statement | 5% (11%) | 30% (25%) | 35% (20%) | 35% (22%) | 17% (15%) |
| Biology Talk Total | 5% (11%) | 31% (26%) | 42% (42%) | 37% (23%) | 17% (15%) |
| How to/ Technical | 4% (9%) | 22% (18%) | 28% (17%) | 27% (19%) | 12% (12%) |
| Social Negotiation | 3% (6%) | 27% (22%) | 23% (17%) | 40% (24%) | 24% (22%) |
| Read Aloud | 4% (9%) | 27% (24%) | 28% (18%) | 26% (18%) | 11% (10%) |
| Refer to Other Time or Place | 0.2% (0.6%) | 0.6% (1%) | 0.3% (1%) | 1% (3%) | |
| Refer to Other Table Activity | | 1% (3%) | 0.9% (3%) | 0.3% (0.9%) | |
| Unintelligible Talk | 1% (3%) | 10% (23%) | 9% (17%) | 6% (14%) | 11% (16%) |
| Enjoy Experience | 1% (2%) | 9% (11%) | 6% (11%) | 12% (14%) | 2% (3%) |
| Dislike/ Frustration | | 2% (4%) | 3% (6%) | 2% (4%) | 0.3% (0.8%) |

Notes: Biology Talk Total includes intervals in which observers coded either a Biology Question, a Biology Statement, or both. Since only a few people were in groups of 6 or 7, we combine these with groups of 5 to better represent these larger groups.
Total Intervals N=1827.

Table 48: Proportion of Engagement Behaviors over Total Available Intervals by Group Size for DeepTree Naturalistic Study

| Behavior | 1 person (n = 48) | 2 people (n = 133) | 3 people (n = 83) | 4 people (n = 24) | 5+ people * (n=28) |
|---|---|---|---|---|---|
| Pull Person/ Interest | 3% (15%) | 1% (8%) | 2% (12%) | 2% (10%) | |
| Move Person / Help | 0.6% (3.6%) | 0.1% (0.7%) | 3% (8%) | 0.8% (2.8%) | 0.2% (0.5%) |
| Prevent Touch/ Control | 0.4% (2.2%) | 6% (16%) | 10% (21%) | 3% (7%) | 6% (6%) |
| Yield | | 0.7% (2.6%) | 0.8% (3%) | | 0.5% (0.8%) |
| 1 Manipulate | 85% (29%) | 96% (16%) | 90% (23%) | 92% (14%) | 92% (11%) |
| Turn Taking | 0.3% (1.9%) | 11% (18%) | 10% (21%) | 11% (19%) | 26% (16%) |
| 2 Manipulate | 5% (11%) | 40% (35%) | 60% (38%) | 57% (33%) | 46% (11%) |
| 3+ Manipulate | 0.3% (1.3%) | 3% (10%) | 24% (33%) | 27% (25%) | 15% (17%) |
| Point/ Indicate/ No touch | 2% (7%) | 13% (18%) | 14% (20%) | 22% (22%) | 28% (17%) |
| Biology Question | 1% (5%) | 16% (21%) | 13% (22%) | 10% (17%) | 0.7% (1%) |
| Biology Statement | 5% (13%) | 32% (32%) | 24% (35%) | 29% (39%) | 17% (15%) |
| Biology Talk | 5% (14%) | 35% (33%) | 27% (36%) | 31% (39%) | 16% (15%) |
| How To/ Technical | 9% (20%) | 35% (32%) | 34% (35%) | 31% (29%) | 12% (12%) |
| Social Negotiation | 7% (18%) | 26% (31%) | 34% (35%) | 29% (30%) | 24% (22%) |
| Read Aloud | 6% (17%) | 23% (29%) | 19% (28%) | 17% (26%) | 11% (10%) |
| Refer to Other Time/ Place | 1% (7%) | 3% (7%) | 2% (7%) | 1% (7%) | |
| Refer to Other Table Activity | | 0% (0.3%) | 0% (0.4%) | 0.8% (2.4%) | |
| Unintelligible Talk | 3% (12%) | 6% (16%) | 9% (23%) | 10% (28%) | 11% (16%) |
| Enjoy Experience | 4% (10%) | 16% (22%) | 18% (29%) | 21% (32%) | 2% (3%) |
| Dislike/Frustration | 0.1% (0.5%) | 2% (6%) | 2% (7%) | 11% (15%) | 0.3% (0.8%) |

Notes: Biology Talk Total includes intervals in which observers coded either a Biology Question, a Biology Statement, or both. Since only a few people were in groups of 6 or 7, we combine these with groups of 5 to better represent these larger groups.
Total Intervals N=1176.

## *BAT Event Proportions*

Table 49: Proportion of Engagement Behaviors over Total Available Intervals by Group Size for BAT Video Study

| Behavior | 1 person (n = 5) | 2 people (n = 12) | 3 people (n = 2) | 4 people (n = 4) |
|---|---|---|---|---|
| Pull Person/ Interest | | | | |
| Move Person / Help | | | | |
| Prevent Touch/ Control | | | | |
| Yield | | | | |
| 1 Manipulate | 97% (4%) | 98% (2%) | 66% (15%) | 90% (1%) |
| Turn Taking | 0.4% (0.8%) | 26% (14%) | 2% (3%) | 21% (1%) |
| 2 Manipulate | 0.4% (0.8%) | 24% (14%) | 2% (3%) | 51% (3%) |
| 3 Manipulate | | | | 15% (1%) |
| Point/ Indicate/ No touch | | 31% (16%) | | 57% (3%) |
| Biology Question | 0.7% (1.6%) | 14% (6%) | 5% (7%) | 10% (1%) |
| Biology Statement | 5% (7%) | 58% (19%) | 14% (20%) | 21% (1%) |
| Biology Talk | 5% (7%) | 62% (17%) | 17% (24%) | 31% (2%) |
| How To/ Technical | 4% (6%) | 22% (10%) | 14% (20%) | 5% (0%) |
| Social Negotiation | 0.7% (1.6%) | 37% (8%) | 2% (3%) | 15% (1%) |
| Read Aloud | 2% (3%) | 42% (14%) | 2% (3%) | 31% (2%) |
| Refer to Other Time/ Place | | 1% (2%) | | |
| Refer to Other Table Activity | | 0.5% (0.8%) | | |
| Unintelligible Talk | | 6% (5%) | | 10% (1%) |
| Enjoy Experience | 4% (8%) | 10% (7%) | | 10% (1%) |
| Dislike/Frustration | 0.4% (0.8%) | 0.7% (1.1%) | | |

Note: Biology Talk Total includes intervals in which observers coded either a Biology Question, a Biology Statement, or both.
Total Intervals N=407.

Table 50: Proportion of Engagement Behaviors over Total Available Intervals by Group Size BAT Naturalistic Study

| Behavior | 1 person (n = 24) | 2 people (n = 50) | 3 people (n = 47) | 4 people (n = 18) | 5 people (n = 6) | 6 people (n = 10) |
|---|---|---|---|---|---|---|
| Pull Person/ Interest | 1% (5%) | 2% (6%) | 4% (13%) | 1% (6%) | | |
| Move Person / Help | 0.9% (4%) | 3% (7%) | 4% (13%) | 3% (8%) | 3% (3%) | 20% (5%) |
| Prevent Touch/ Control | 2% (5%) | 9% (15%) | 11% (18%) | 3% (10%) | 17% (9%) | 10% (13%) |
| Yield | 0.3% (2%) | 1% (4%) | 0.2% (0.8%) | | | 3% (5%) |
| 1 Manipulate | 91% (24%) | 98% (7%) | 97% (12%) | 76% (39%) | 100% (0%) | 100% (0%) |
| Turn Taking | 1% (3%) | 13% (16%) | 7% (11%) | 7% (9%) | 18% (10%) | 9% (8%) |
| 2 Manipulate | 6% (14%) | 44% (34%) | 53% (33%) | 52% (30%) | 43% (22%) | 91% (10%) |
| 3 Manipulate | 0.1% (0.3%) | 1% (5%) | 16% (30%) | 4% (13%) | 6% (3%) | 33% (31%) |
| Point/ Indicate | 4% (9%) | 16% (18%) | 12% (16%) | 11% (20%) | 41% (24%) | 26% (5%) |
| Biology Question | 3% (6%) | 18% (23%) | 14% (22%) | 10% (15%) | 26% (13%) | 34% (18%) |
| Biology Statement | 9% (17%) | 33% (28%) | 32% (35%) | 33% (40%) | 93% (8%) | 41% (16%) |
| Biology Talk | 9% (18%) | 36% (32%) | 35% (35%) | 35% (39%) | 93% (8%) | 44% (18%) |
| How To/ Technical | 8% (15%) | 34% (23%) | 43% (34%) | 46% (37%) | 44% (23%) | 65% (16%) |
| Social Negotiation | 9% (22%) | 18% (15%) | 21% (24%) | 18% (27%) | 28% (15%) | 55% (20%) |
| Read Aloud | 4% (9%) | 27% (26%) | 27% (30%) | 27% (30%) | 88% (13%) | 53% (14%) |
| Refer to Other Time/ Place | 0.1% (0.3%) | 2% (5%) | 1% (3%) | 3% (8%) | 3% (3%) | |
| Refer to Other Table Activity | | | | | | |
| Unintelligible Talk | 2% (12%) | 9% (21%) | 12% (27%) | | | 7% (7%) |
| Enjoy Experience | 2% (7%) | 7% (11%) | 12% (17%) | 6% (12%) | 22% (11%) | 9% (8%) |
| Dislike/Frustration | 0.6% (2%) | 6% (13%) | 2% (8%) | 2% (4%) | 17% (41%) | |

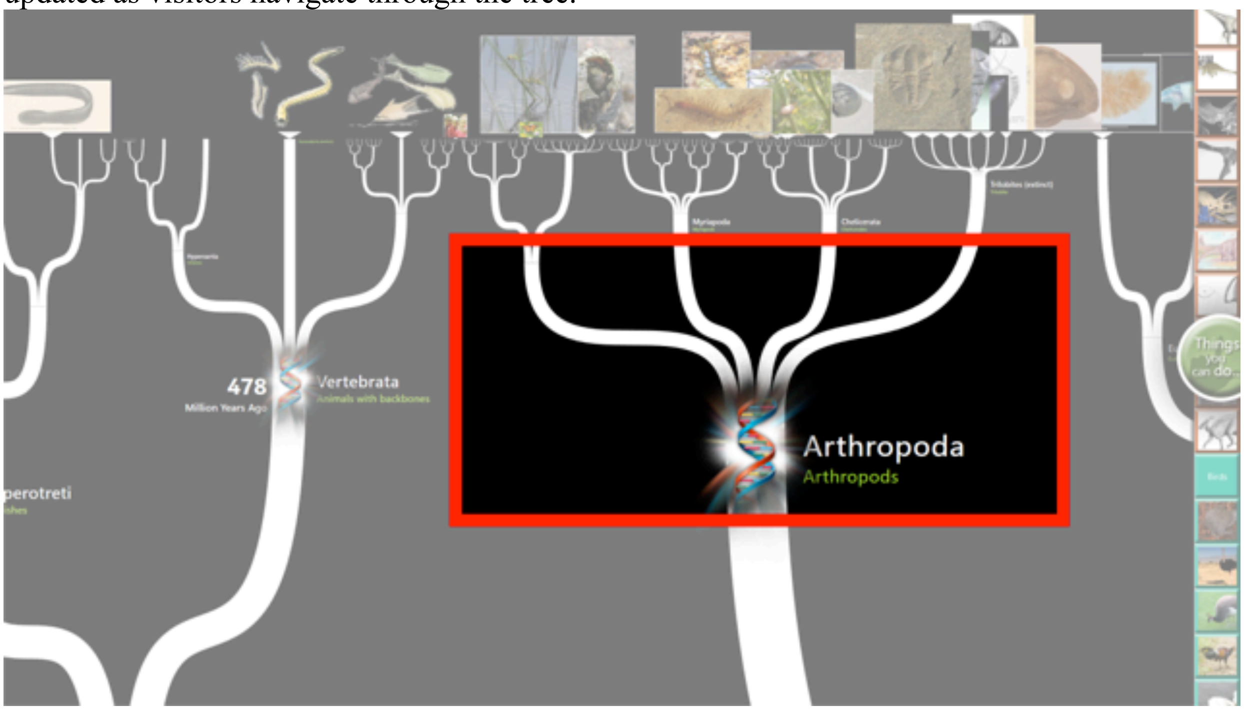
Note: Biology Talk Total includes intervals in which observers coded either a Biology Question, a Biology Statement, or both.
Total Intervals N=689; Total Visitors N=155

## Table Log Summary Measures Definitions

# Semantic Logging Measure Guide

## Definitions

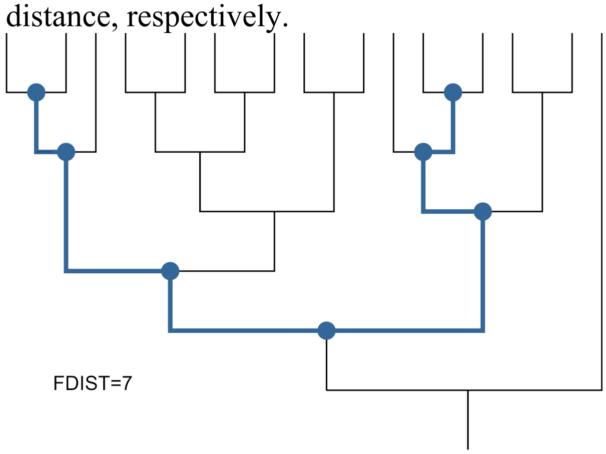
- Focal Node: Internal node that fills up most screen space. Focal node is continuously updated as visitors navigate through the tree.



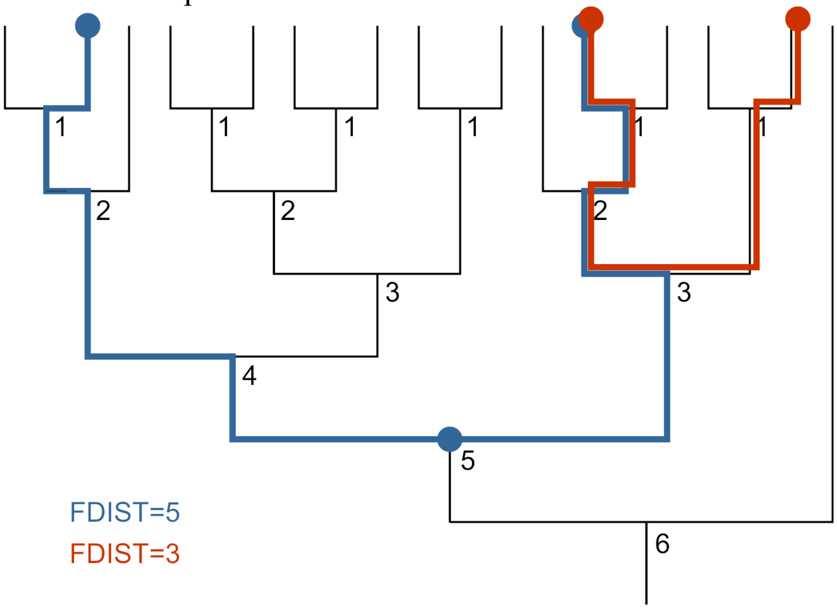

## Unit Definition

The following units are defined (every variable name will contain the unit for reference):

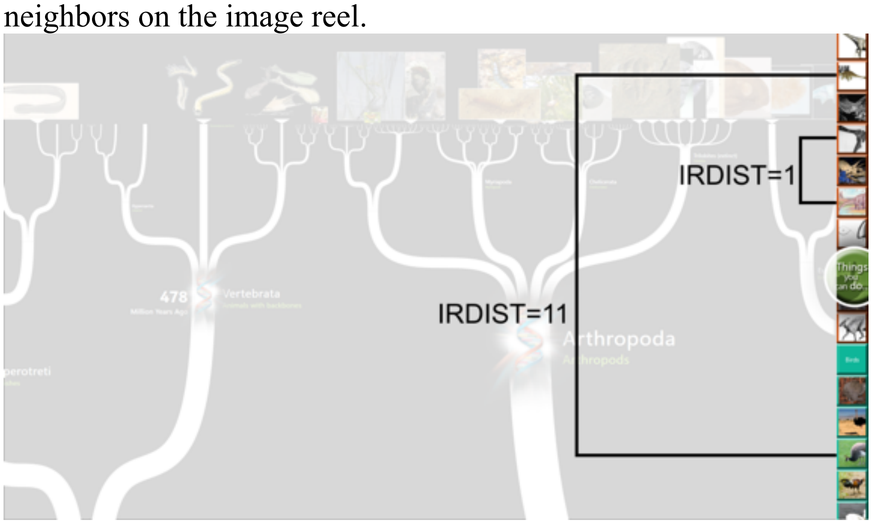
- Flight-distance (FDIST): Distance in number of focal nodes that separate one point of the tree from another. This is direct proportional to flight-time, and experienced flight-distance, respectively.




- Relatedness-distance (RDIST): For any two species A and B, RDIST specifies the time sequence index of the MRCA of A and B. The time sequence index is assigned to internal nodes so that the root has the highest index, and any children always has a lower index than its parent node.




- Screen-distance (SDIST): Distance in screen pixels.
- Image reel distance (IRDIST): For any two image reel items A and B, IRDIST is the number of image reel items between A and B. A value of zero indicates that A and B are neighbors on the image reel.




- Time in seconds (TIME): Time in seconds.
- Number of occurrences (COUNT): Number a certain element has occurred during usage.
- Scale (SCALE): Decimal scale (see definition of each variable).
- Average (AVRG_<UNIT>): Average of the given set in the defined unit.
- Standard deviation (STDD_<UNIT>): Standard deviation of the given set in the defined unit. Log Measures

## Log Measures

**General (GEN)**

- GEN_TIME_SPENT: Time spent at the exhibit.
- GEN_COUNT_RETURN: Number of times people selected return to root.
- GEN_COUNT_ATTRACTOR _UNIQUE_FOCAL_NODES: Number of unique focal nodes that appear during the automatic zoom/ attractor display while people are at the table. We think this is a very passive view of the tree, but still more important is to be able to exclude these from the overall count of unique focal nodes viewed during more active navigation.

**Inspection (INS)**

- INS_COUNT_TEXT_10_SEC: Number of species text (does it make sense to count species that only have a name separately from those that also have a purple text box?) that are visible on screen for at least 10 seconds. We're trying to exclude fly through text so as to count text that was available for reading, whether or not people read it.
- INS_TIME_TEXT_10_SEC: Total time spent with text visible for at least 10 seconds.
- INS_COUNT_UNIQUE_IMAGE_ZOOMED: Number of unique top images that are zoomed. (We put this under NAV generally rather than Top Image because it can occur during any type of navigation.)
- INS_SCALE_IMAGE_ZOOMED: Total amount zoomed on all images combined. Per top image, value ranges from 1.0 (no zoom), to 3.0 (max zoom).

**Navigation Measures (NAV):**

- NAV_TIME_NAVIGATING: Amount of time spent using any of the navigation types.
- NAV_COUNT_UNIQUE_FOCAL_NODES: Total number of unique focal nodes visited.
- NAV_FDIST: Total number of focal node changes during any of the navigation types.
- NAV_TIME _SIMULTANOUS_*[N]*: Time in which more than *N* navigation action (Top Image Hold, Reel Item Hold, Manual Navigation) is active (N = 1, 2, 3, 4, 5).
- NAV_COUNT _SIMULTANOUS_*[N]*: Number of times in which more than *N* navigation actions (Top Image Hold, Reel Item Hold, Manual Navigation) are active (N = 1, 2, 3, 4, 5).
- NAV_TIME_NAV_WHILE_FLYING: Time in which navigation is attempted (Top Image Hold, Reel Item Hold, Manual Navigation) during an automated fly-through (Find, Relate or Return).
- NAV_COUNT_NAV_WHILE_FLYING: Number of times navigation is attempted (Top Image Hold, Reel Item Hold, Manual Navigation) during an automated fly-through (Find, Relate or Return).

***Manual Navigation (M):***

- NAV_M_TIME_SPENT: Time spent manually navigating.
- NAV_M_FDIST: Total number of focal node changes while manually navigating.
- NAV_M_COUNT_UNIQUE_FOCAL_NODES: Total number of unique focal nodes visited during manual navigation.
- NAV_M_COUNT_INWARD_MOVES;
- NAV_M_COUNT_OUTWARD_MOVES;
- NAV_M_COUNT_HORIZONTAL_MOVES: Total number of inward/ outward/ horizontal manual navigation moves.

***Top Image Navigation (T):***

- NAV_T_TIME_SPENT: Number of seconds spent holding a top image.
- NAV_T_FDIST: Number of focal node changes while a top image is held.
- NAV_T_COUNT_TOTAL_HELD: Total number of times a top image has been held.
- NAV_T_AVRG_TIME_HELD: Average time a top image was held.
- Do you also want a measure of variability, if you're not gathering a list of times? NAV_T_STDD_TIME_HELD? We expect the times to be skewed right so Average isn't necessarily a good measure. Median, or the Distribution would be better.
- NAV_T_COUNT_SPECIES_HELD: Number of unique images held at least once.
- NAV_T_COUNT_SPECIES_NAVIGATED_TO: Number of unique species that became visible while corresponding top image was held.

***Reel Item Navigation (R):***

- NAV_R_COUNT_SPECIES_PULLED_OUT: Number of species pulled out from the reel during free exploration.
- NAV_R_TIME_SPENT: Number of seconds spent holding an image reel item.
- NAV_R_COUNT_TOTAL_HELD: Total number of times a reel item has been held.
- NAV_R_AVRG_TIME_HELD: Average time a reel item was held.
- NAV_R_STDD_TIME_HELD? We expect the times to be skewed right so Average isn't necessarily a good measure. Median, or the Distribution would be better.
- NAV_R_FDIST: Number of focal node changes while a reel item is held.
- NAV_R_COUNT_SPECIES_HELD: Number of unique reel items held at least once.
- NAV_R_COUNT_SPECIES_NAVIGATED_TO: Number of unique species that became visible while corresponding reel item was held.

***Find Navigation (F):***

- NAV_F_TIME_SPENT: Time spent flying after Find.
- NAV_F_FDIST: Total number of focal node changes while Find is active.

***Relate Navigation (REL):***

- NAV_REL_TIME_SPENT: Time spent flying to an MRCA.
- NAV_REL_FDIST: Total number of focal node changes while Relate is flying to the MRCA.

**Relate (REL)**

- REL_TIME_SPENT_SELECTING_SPECIES: Time spent selecting species while the dialog is active.
- REL_COUNT_SPECIES_PULLED_OUT: Number of species pulled out from the reel while selecting species.
- REL_SDIST_SCROLLED: Distance scrolled on image reel while selecting species.
- REL_COUNT_SPECIES_RECENTLY_USED: Number of species selected from "Recently used".
- REL_COUNT_SPECIES_IMAGE_REEL: Number of species selected from the image reel.
- REL_COUNT: Number of relate queries executed.
- REL_COUNT_BOTH_RECENTLY_USED: Number of relate queries with both species selected from "Recently used".
- REL_COUNT_ONE_RECENTLY_USED: Number of relate queries with one species selected from "Recently used".
- REL_COUNT _BOTH_IMAGE_REEL: Number of relate queries with both species selected from the reel.
- REL_AVRG_RDIST_RELATEDNESS: Average relatedness distances of all relate queries.
- REL_STDD_RDIST_RELATEDNESS: Standard deviation of relatedness distances of all relate queries.
- REL_AVRG_IRDIST_REEL_PROXIMITY: Average reel distances of all relate queries where both species are picked from the reel.
- REL_STDD_IRDIST_REEL_PROXIMITY: Standard deviation of reel distances of all relate queries where both species are picked from the reel.
- REL_TIME_SPENT_PICKING_SPECIES: Total time spent picking species.
- REL_TIME_AVERAGE_TIME_SPENT_PICKING_SPECIES: Average time spent picking species per relate query.
- REL_TIME_STDD_TIME_SPENT_PICKING_SPECIES: Standard deviation for time spent picking species per relate query.

**Training Tree (TT)**

- TT_TRAIT_COUNT: Number of traits selected in the Training Tree across all Trait Displays.
- TT_ COUNT_UNIQUE_TRAITS: Number of unique traits selected in the Training Tree across all Trait Displays.
- TT_ AVRG_TIME_SPENT_PER_TRAIT: Average time spent per trait across all Trait Displays.
- Again, if collecting averages, then variability may also be useful, i.e., TT_STDD_TIME_SPENT_PER_TRAIT. But these two may be lower priority for us.

**Trait Display (TRAIT)**

- TRAIT_TIME_SPENT: Time spent in the Trait Display.
- TRAIT_COUNT: Number of times that the Trait display is activated
- TRAIT_TIME_SPENT_FROM_RELATE: Time spent in the Trait Display when coming from a relate query.
- TRAIT_COUNT_FROM_RELATE: Number of times the Trait Display was entered via Relate.
- TRAIT_TIME_SPENT_WHILE_BROWSING: Time spent in the Trait Display while browsing.
- TRAIT_COUNT_WHILE_BROWSING: Number of times the Trait Display was entered via the glowing double-helix while browsing.
- TRAIT_UNIQUE_COUNT: Number of unique nodes/ trait displays activated
- TRAIT_COUNT_UNIQUE_ILLUSTRATIONS_SELECTED: Number of trait illustrations (unique trait illustrations?) highlighted (selected/ enlarged other than the default one) across all Trait Displays
- TRAIT_AVRG_ILLUSTRATIONS_SELECTED
- TRAIT_STDD_ILLUSTRATIONS_SELECTED: Average and Standard deviation of the number of illustrations selected per Trait Displayed. (This is lower priority.)

**FloTree (FT)**

- FT_COUNT: Number of times FloTree/ Experiment was activated.
- FT_TIME: Time spent in the FloTree
- FT_COUNT_EXPERIMENTS: Total number of experiments launched.
- FT_COUNT_SPECIATION_EXPERIMENTS: Number of experiments with at least one speciation event
- FT_COUNT_SPECIES: Total number of species created in experiments. (Maybe also list of number of species created so we can get the distribution of these…but that may be lower priority. However, we wonder whether there's a minimum number of species created (>2) that actually makes a difference in people's perceptions.)
- FT_AVRG_SPECIES: Average number of species created per experiment.
- FT_STDD_SPECIES: Standard deviation of number of species created per experiment – would serve as a proxy if we don't have list of number of species created.)
- FT_COUNT_TEXT_EXPLANATIONS: Number of times text explanations are activated during FloTree.
- FT_COUNT_TREE_DIAGRAM: Number of times they transform population view to tree view.

## Summary Statistics for Table Log Measures

The following tables provide descriptive statistics for the complete set of Table Log Measures.

Table 51: General Table Activity Descriptives

| | Mean (SD) | Median | Min-Max |
|---|---|---|---|
| General Time Spent | 529.56 s. | 455 | 15-3259 s. |
| Gen Count Return | .53 (1.95) | 0 | 0-19 |

Table 52: Navigation Activity Descriptives

| | Mean (SD) | Median | Min-Max |
|---|---|---|---|
| Nav Count While Flying * | 2.88 (19.38) | 0 | 0-179 |
| Nav Count Simultaneous 1 | 31.48 (61.19) | 5 | 0-412 |
| Nav Count Simultaneous 2 | 9.94 (42.70) | 0 | 0-348 |
| Nav Count Simultaneous 3 | 10.08 (34.66) | | 0-280 |
| Nav Count Simultaneous 4 | 5.36 (21.42) | 0 | 0-146 |
| Nav Count Simultaneous 5 | 2.90 (15.93) | 0 | 0-100 |
| Nav Time Nav While Flying | .90 (4.20) | 0 | 0-28.91 |
| Nav Time Navigating | 116.79 (176.39) | 57.19 | 0-1168.17 |
| Nav Time Simultaneous 1 | 33.80 (79.54) | 1.35 | 0-489.73 |
| Nav Time Simultaneous 2 | 6.64 (29.37) | 0 | 0-265.79 |
| Nav Time Simultaneous 3 | 20.90 (76.50) | 0 | 0-512.02 |
| Nav Time Simultaneous 4 | 4.72 (20.19) | 0 | 0-131.72 |
| Nav Time Simultaneous 5 | 1.49 (7.94) | 0 | 0-53.47 |
| Nav Total Held | 2.31 (3.89) | 0 | 0-15 |
| Nav Count Unique Foc Nodes | 118.09 (79.82) | 117 | 0-368 |
| Nav F Dist | 369.62 (405.52) | 290 | 0-3755 |
| Nav F Time Spent | No data | 0 | |
| Nav M Horizontal Moves | 30.59 (30.68) | 25 | 0-207 |
| Nav M Count Inward Moves | 17.41 (23.65) | 11 | 0-136 |
| Nav M Outward Moves | 32.85 (41.26) | 16 | 0-203 |
| Nav M Count Unique Foc Nodes | 32.38 (41.28) | 12 | 0-151 |
| Nav M FDist | 75.85 (140.29) | 31 | 0-758 |
| Nav M Time Spent | 81.75 (107.64) | 53 | 0-668 |
| Nav R Avg Time Held | 2.70 (2.49) | 2.40 | 0-10.36 |
| Nav R Count Spec Held | 5.09 (6.07) | 3 | 0-25 |
| Nav R Count Spec Navigated To | 2.33 (3.87) | 0 | 0-15 |
| Nav R Spec Pulled Out | 5.36 (6.92) | 3 | 0-31 |

| | | | |
|---|---|---|---|
| Nav R FDist | 356.22 (1231.08) | 45 | 0-9249 |
| Nav R STDD Time Held | 2.70 (2.49) | 2.40 | 0-10.36 |
| Nav R Time Spent | 89.85 (124.86) | 19.37 | 0-535.77 |
| Nav Rel FDist | 59.04 (141.07) | 0 | 0-1087 |
| Nav Rel Time Spent | 20.52 (47.14) | 0 | 0-361.57 |
| Nav T Avg Time Held | 1.02 (1.01) | .93 | 0-5.87 |
| Nav T Count Spec Held | 4.82 (6.95) | 3 | 0-48 |
| Nav T Count Spec Nav To | 1.06 (1.45) | 0 | 0-6 |
| Nav T Count Total Held | 1.04 (1.41) | 0 | 0-6 |
| Nav T FDist | 201.44 (1549.01) | 4 | 0-14205 |
| Nav T STDD Time Held | .37 (.50) | .14 | 0-2 |
| Nav T Time Spent | 15.00 (20.10) | 5.93 | 0-132.96 |

Note: *Count of number of simultaneous touches while flying. A measure of frequency of multiple users interacting with the table at once.

Table 53: Inspection Activity Descriptives

| | Mean (SD) | Median | Min-Max |
|---|---|---|---|
| INS Scale Image Zoomed | No data | | |
| INS Count Unique Image Zoomed | No data | 0 | |
| INS Count Text 10 Sec | 4.8 (5.07) | 3 | 0 -27 |
| INS Time Text 10 sec | 502.23 (1460.21) | 101.31 | 0 – 12076.01 |

Table 54: Trait Activity Descriptives

| | Mean (SD) | Median | Min-Max |
|---|---|---|---|
| Trait Time Spent | 60.23 (103.52) | 23.97 | 0-738.13 |
| Trait Unique Count | 1.58 (3.36) | 0 | 0-16 |
| Trait Count | 2.45 (2.92) | | 0-20 |
| Trait Unique Illustrations | 2.17 (4.18) | 1 | 0-19 |
| Trait STDD Illustrations | .40 (.68) | 0 | 0-3.76 |
| Trait Avg Illustrations | 2.38 (8.13) | 2 | 0-48 |

Table 55: Relate Function Descriptives

| | All Visitors (N=169) | | | Visitors who Found Relate (N=54) | | |
|---|---|---|---|---|---|---|
| **Measure** | **Mean (SD)** | **Median** | **Min-Max** | **Mean (SD)** | **Median** | **Min-Max** |
| Rel Avg IRDist Reel Proximity | 7.58 (17.59) | 0 | 0-100 | 23.73 (24.29) | 18.50 | 1-100 |
| Rel Avg Dist Relatedness | 19.53 (34.87) | 0 | 0-128 | 61.11 (35.58) | 66.50 | 14-128 |
| Rel Count | .93 (2.11) | 0 | 0-17 | 2.91 (2.88) | 2 | 1-17 |
| Rel Count Both Image Reel | .72 (2.01) | 0 | 0-17 | 2.24 (3.06) | 1 | 0-17 |
| Rel Count Both Recent Used | .03 (.17) | 0 | 0-1 | .09 (.29) | 0 | 0-1 |
| Rel Count One Recent Used | .18 (.48) | 0 | 0-2 | .57 (.72) | 0 | 0-2 |
| Rel Count Spec Image Reel | 1.62 (4.06) | 0 | 0-34 | 5.06 (5.89) | 3 | 1-34 |
| Rel Count Spec Pulled Out | 1.75 (4.49) | 0 | 0-39 | 5.48 (6.55) | 4 | 0-39 |
| Rel Count Spec Rec Used | .24 (.57) | 0 | 0-2 | .76 (.80) | 1 | 0-2 |
| Rel SDist Scroll | 17.44 (288.95) | 0 | -863-1605 | 300.54 (416.65) | 127 | 0-1605 |
| Rel STDD RDist Proximity | 1.98 (6.44) | 0 | 0-34.29 | 6.21 (10.24) | 1.85 | 0-34.29 |
| Rel STDD RDist Relatedness | 3.16 (8.36) | 0 | 0-42.78 | 9.79 (12.41) | 5.52 | 0-42.78 |
| Rel Time Spent Select Spec | 44.73 (69.87) | 9.33 | 0-386.06 | 71.16 (59.62) | 67.89 | 8.76 – 386.06 |
| Nav Rel FDist | 59.04 (141.07) | 0 | 0-1087 | 184.76 (198.53) | 117.5 | 3-1087 |
| Nav Rel Time Spent | 20.52 (47.14) | 0 | 0-361.57 | 64.23 (64.77) | 40.29 | 5.91-361.57 |

Table 56: Training Tree Activity Descriptives

| | All Visitors (N=169) | | | Visitors who Found Relate (N=54) | | |
|---|---|---|---|---|---|---|
| **Measure** | **Mean (SD)** | **Median** | **Min-Max** | **Mean (SD)** | **Median** | **Min-Max** |
| TT Avrg Time Spent per Trait | 2.63 (6.20) | 0 | 0 – 34.18 | 8.22 (8.66) | 7.30 | 0 – 34.18 |
| TT STDD Time Spent per Trait | .63 (2.43) | 0 | 0 – 19.54 | 1.98 (4.00) | .29 | 0-19.55 |
| TT Trait Count | 3.21 (9.80) | 0 | 0 -80 | 10.04 (15.31) | 5.50 | 0-80 |
| TT Unique Trait Count | 1.50 (3.38) | 0 | 0 – 16 | 4.69 (4.57) | 5.50 | 0-16 |

Table 57: FloTree Descriptives

| | All Visitors (N=169) | | | Visitors who Found FloTree (N=89) | | |
|---|---|---|---|---|---|---|
| | Mean (SD) | Median | Min-Max | Mean (SD) | Median | Min-Max |
| Time in FloTree | 89.32 s. | 53.23 | 0 – 585.66 s. | 169.61 (97.04) | 155.47 | 10.56-585.66 |
| FT Count | .75 (.80) | 1 | 0 – 5 | 1.42 (0.75) | 1 | 1-5 |
| FT Count Experiments | .98 (1.27) | 0 | 0-8 | 1.87 (1.18) | 2 | 0-8 |
| FT Count Speciation Experiments | .82 (1.12) | 0 | 0-6 | 1.56 (1.11) | 1 | 0-6 |
| FT Count Species | 4.48 (6.79) | 0 | 0 – 40 | 8.51 (7.28) | 7 | 0-40 |
| FT Count Tree Diagram | 5.27 (8.12) | 1 | 0-56 | 10.01 (8.82) | 7 | 0-56 |
| FT Avrg Species | 2.20 (2.66) | 0 | 0 – 9.33 | 4.19 (2.26) | 4 | 0-9.33 |
| FT STDD Species | .30 (.54) | 0 | 0 – 2.47 | 0.57 (0.63) | 0.35 | 0-2.47 |

## Confirmatory Factor Analysis of Exhibit Engagement

Our attempt at creating an overall measure of exhibit engagement went through the following steps. Through a process of theory-based data reduction — the evaluation team rated the importance of *all* the table measures, with the goal of selecting 12-20 that would capture the most important features of visitor interactions with the exhibit. Sixteen (16) variables were identified from the table log data covering all of the important exhibit activity features (Table 58). It was hypothesized that each of these measures would contribute in some fashion to an assessment of how engaged a person was with the exhibit. By specifying the relationships of these indicators on a latent factor of engagement, confirmatory factor analysis would allow us to assess the overall fit of this engagement model and, if appropriate, to export a latent factor score for each person's engagement. This new score could be used both as an outcome variable to be predicted by person characteristics (e.g. age, education, group size) and as a predictor for learning from the exhibit.

Table 58: List of Indicators for Confirmatory Factor Analysis

| Short Name | Indicator |
|---|---|
| GenTime | GEN_TIME_SPENT: Time spent at the exhibit. |
| Text10 | INS_COUNT_TEXT_10_SEC: Number of species text that are visible on screen for at least 10 seconds. |
| FocNode | NAV_COUNT_UNIQUE_FOCAL_NODES: Total number of unique focal nodes visited |
| NavTime | NAV_TIME_NAVIGATING: Amount of time spent using any of the navigation types. |
| RNav | NAV_R_COUNT_SPECIES_NAVIGATED_TO: Number of unique species that became visible while corresponding reel item was held. |
| TNav | NAV_T_COUNT_SPECIES_HELD: Number of unique images held at least once. |
| Relate | REL_COUNT: Number of relate queries executed |
| RelPull | REL_COUNT_SPECIES_PULLED_OUT: Number of species pulled out from the reel while selecting species. |
| TraitNum | TT_TRAIT_COUNT: Number of traits selected in the Training Tree across all Trait Displays |
| FTSpec | FT_COUNT_SPECIATION_EXPERIMENTS: Number of experiments with at least one speciation event |
| Comp | Sum of the binary coding of FTSpec, RNav, TNav, Relate, RelPull, and TraitNum |

Prior to fitting the factor models, the distribution of each variable was examined and, where appropriate, log transformed to meet the assumptions of normality (i.e. to reduce skew). Several of the measures did not carry enough variability to be considered a continuous variable (FTSpec, RNav, TNav, Relate, RelPull, TraitNum), making them unusable in a confirmatory factor analysis (CFA). Therefore, a composite score was created such that each of these indicators was first made into a binary variable—a "1" score indicated that the individual experienced this aspect of the table, whereas a "0" indicated that they did not. Each person was then given a summed composite score to represent how many of these exhibit features they experienced.

We used LISREL software (Jöreskog & Sörbom, 2012) to fit the CFA models. **Error! Reference source not found.** shows the structure of the best model fit and the resultant factor loadings. The subsequent tables show indicator reliabilities and indicator-factor correlations.

Figure 21: CFA – Table Engagement: Factor Loadings

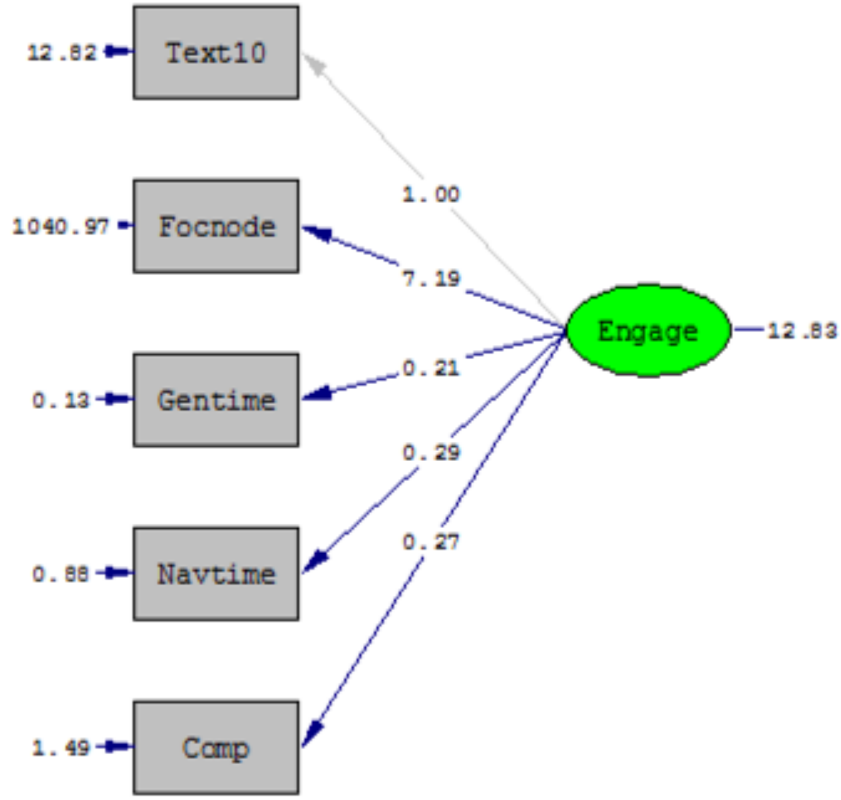



Table 59: LOE Indicator Reliabilities (Item Reliability Estimates)

| | | Estimated Indicator Parameters | | |
|---|---|---|---|---|
| **Indicator** | **Observed Variance** | **Error Variance** | **True Variance** | **Reliability** |
| Text10 | 25.650 | 12.820 | 12.830 | .500 |
| FocNode | 1703.700 | 1040.965 | 662.735 | .389 |
| Comp | 2.439 | 1.493 | .946 | .388 |
| Gentime | .690 | .131 | .559 | .810 |
| Navtime | 1.962 | .877 | 1.085 | .553 |

Note: Reliability = True/ Observed

Table 60: Indicator-factor Correlation (Validity Estimates)

| | | | Estimates | | |
|---|---|---|---|---|---|
| **Indicator** | **Observed Indicator Variance** | **Factor** | **Factor Variance** | **Factor Loading** | **Indicator-factor Correlation** |
| Text10 | 25.650 | Engage | 12.83 | 1.000 | .707 |
| FocNode | 1703.700 | | | 7.187 | .624 |
| Comp | 2.439 | | | 0.272 | .623 |
| Gentime | .690 | | | 0.209 | .901 |
| Navtime | 1.962 | | | 0.291 | .744 |

Note: Indicator– Factor correlation: Loading*(√(Factor variance/ Indicator variance)

*Model Fit*

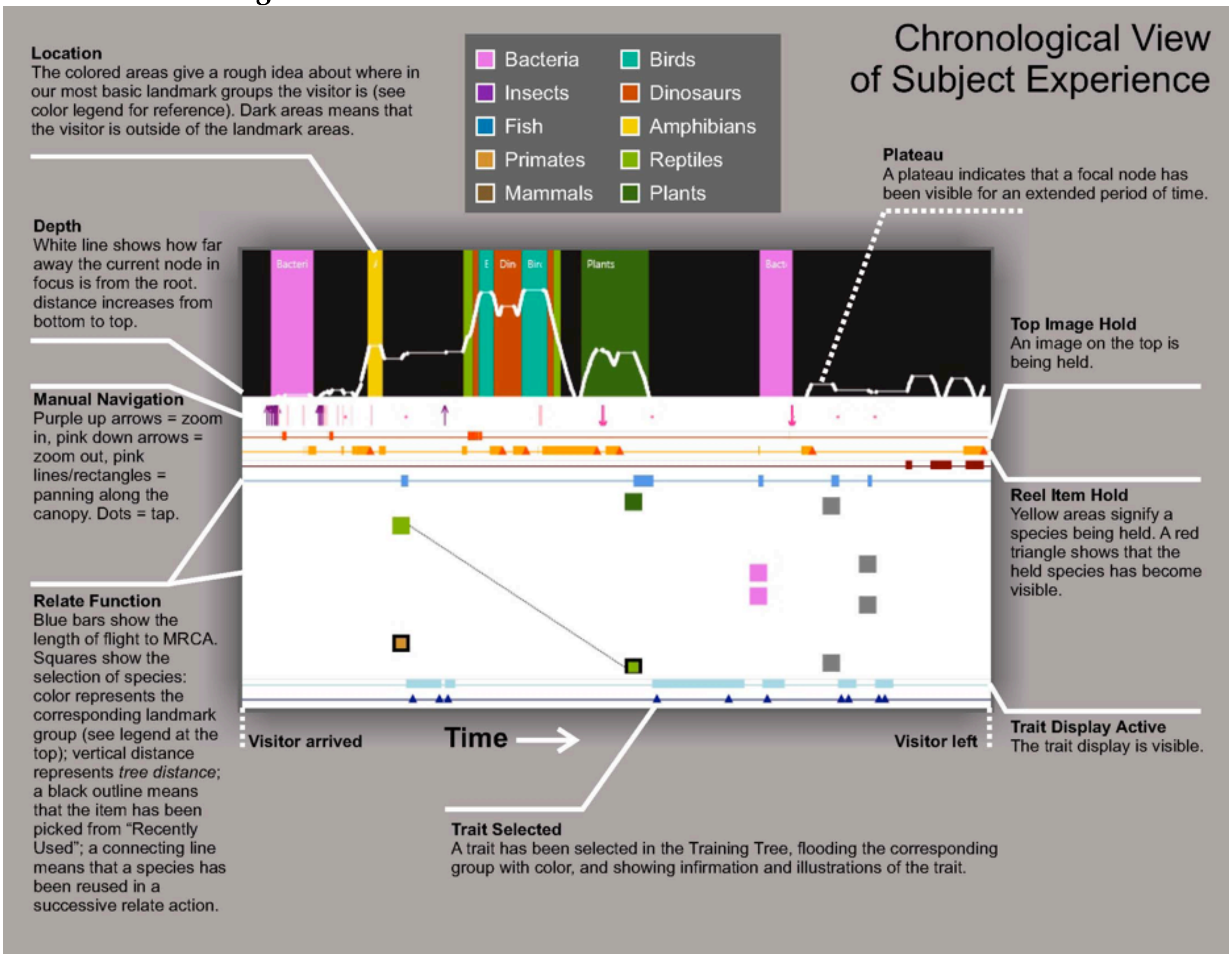
Model fit statistics were examined for the one-factor model of engagement described above. The Minimum Fit Function Chi-Square suggests a poor overall fit to the data (*df* = 5, $\chi^2 = 31{,}697$, $p < .001$). The Root Mean Square Root of Approximation (*RMSEA* = .174, 90% *CI* = 0.117, .235), *RMR* (4.10) and the *CFI* (.941) also lend evidence to suggest poor model fit.

*Conclusion*

It was determined by the evaluation team that the CFA results would not be reliable enough to convert into latent factor scores for each individual. Due to overall poor model fit and low reliability in several of the indicator measures, it was determined that more work would be necessary to determine a better fitting model before utilizing the model to create composite engagement scores.

An encouraging aspect of the factor analysis is that general time spent was the most reliable and valid indicator of engagement. Therefore, we believe that it is reasonable to use time spent at the exhibit as a proxy for engagement, since a more complex measure was found to be premature.

## Semantic Log Visualization Guide



## Survey Text

| ID #: |
|---|

Date and Time______________________

Life on Earth Survey

1. About how many minutes did you spend at the touch table? (please circle one)

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 More: _____

2. What was the touch table about?

3. Please circle your answer.

| | | | | | | |
|---|---|---|---|---|---|---|
| a. How interesting was the touch table? | Not interesting<br>1 | 2 | Some<br>3 | 4 | Very interesting<br>5 | |
| b. How much did you like using the touch table? | Did not like<br>1 | 2 | Some<br>3 | 4 | Liked a lot<br>5 | |
| c. How much did you learn at the touch table? | Learned nothing<br>1 | 2 | Some<br>3 | 4 | Learned a lot<br>5 | |
| d. How did others at the table affect your learning? | Much harder<br>1 | 2 | Did not affect<br>3 | 4 | Much easier<br>5 | I was alone at table |
| e. How much did you like working with others at the table? | Did not like<br>1 | 2 | Some<br>3 | 4 | Liked a lot<br>5 | I was alone at table |
| f. Would you tell a friend to visit the table? | No<br>1 | 2 | Maybe<br>3 | 4 | Yes<br>5 | |

TURN TO BACK OF PAGE, PLEASE→

4. Your “ancestors” are your parents, grandparents, great-grandparents, and so on. Here are some ideas about the ancestors of different living things.

How much do you agree or disagree with these ideas? *Please circle your answer.*

| | Disagree a lot | Disagree some | Neither | Agree some | Agree a lot |
|---|---|---|---|---|---|
| a. RABBITS and LIZARDS had the same ancestor a long, long time ago. | 1 | 2 | 3 | 4 | 5 |
| b. HUMANS and MUSHROOMS had the same ancestor a long, long time ago. | 1 | 2 | 3 | 4 | 5 |
| c. MICE and RATS had the same ancestor a long, long time ago. | 1 | 2 | 3 | 4 | 5 |
| d. BEARS and SUNFLOWERS had the same ancestor a long, long time ago. | 1 | 2 | 3 | 4 | 5 |

5. Here are some ideas about evolution.
How much do you agree or disagree with these ideas? *Please circle your answer.*

| | Disagree a lot | Disagree some | Neither | Agree some | Agree a lot |
|---|---|---|---|---|---|
| a. CARDINALS, a type of bird, are changing over time. They might be VERY different millions of years in the future. | 1 | 2 | 3 | 4 | 5 |
| b. Most living things today are VERY different from their ancestors who lived a long, long time ago. | 1 | 2 | 3 | 4 | 5 |
| c. HUMAN BEINGS, a type of primate, are changing over time. They might be VERY different millions of years in the future. | 1 | 2 | 3 | 4 | 5 |
| d. Evolution is still going on TODAY. | 1 | 2 | 3 | 4 | 5 |
| e. COYOTES, a type of mammal, are changing over time. They might be VERY different millions of years in the future. | 1 | 2 | 3 | 4 | 5 |

6. Is there anything else you want to tell us about your experience at the table?

TURN TO BACK OF PAGE, PLEASE→

## About You

1. About how many times have you been to a museum in the last year? ___________
2. Have you used a touch table before today? Yes No

| 3. How much have you used iPads, smart phones, or other touch devices? | Not at all<br>1 | 2 | Some<br>3 | 4 | A lot<br>5 |
|---|---|---|---|---|---|

4. Gender (check one): ☐ Male/ Boy ☐ Female / Girl
5. Age (check one)

☐ Younger than 9 years
☐ 9-14 years
☐ 15-17 years
☐ 18-23 years
☐ 24-29 years
☐ 30-39 years
☐ 40-49 years
☐ 50-59 years
☐ 60-69 years
☐ 70+ years

6. What is your home state or country? ______________________________________
7. What is your zip code? _________________________
8. What languages do you speak at home?

☐ English ☐ Spanish ☐ Chinese ☐ Tagalog ☐ Vietnamese

☐ Other: _____________________________________________________________

9. Are you Hispanic or Latino/a? Yes No
10. Race (check all that apply):

☐ Asian American
☐ Native American or Alaskan Native
☐ White or Caucasian
☐ Black or African American
☐ Pacific Islander
☐ Other: _________________

11. How much school have you finished?
*Check your highest completed education level*:

☐ Some elementary school
☐ Some middle school
☐ Some high school
☐ High school diploma/GED
☐ Some college
☐ 2-yr college degree
☐ 4-yr college degree (BA, BS)
☐ Master's degree
☐ Doctoral degree
☐ Professional degree (MD, JD)
☐ Other/different system
____________________________

12. If you have completed a college degree or higher, is your degree biology related? Yes No

# Survey Results

## *DeepTree Survey Results*

Table 61: Self-reported ratings of exhibit experience, DeepTree Video Study

| Question | N | Mean (SD) | Median |
|---|---|---|---|
| How interesting was the touch table? | 131 | 4.13 (0.94) | 4 |
| How much did you like using the touch table? | 129 | 4.10 (0.94) | 4 |
| How much did you learn at the touch table? | 130 | 3.37 (1.16) | 3 |
| How did others at the table affect your learning? | 124 | 3.28 (1.27) | 3 |
| How much did you like working with others at the table? | 123 | 3.54 (1.23) | 4 |
| Would you tell a friend to visit the table? | 131 | 3.98 (1.22) | 4 |

Note: Ratings on a 5 point scale. N=132.

Table 62: Levels of Agreement with Survey Questions about Common Ancestry of Various Species. DeepTree Video Study

| Species Pairs | Mean (SD) | Median | Min- Max |
|---|---|---|---|
| Rabbits and Lizards | 3.88 (1.16) | 4 | 1-5 |
| Humans and Mushrooms | 3.23 (1.53) | 3 | 1-5 |
| Mice and Rats | 4.51 (0.85) | 5 | 1-5 |
| Bears and Sunflowers | 3.27 (1.54) | 3 | 1-5 |
| **Common Ancestor Composite** | **3.71 (1.05)** | **3.75** | **1-5** |

Note: Ratings on a 5 point scale. N=125

Table 63: Levels of Agreement with Survey Questions about Evolution. DeepTree Video Study

| Statement | Mean (SD) | Median | Min- Max |
|---|---|---|---|
| CARDINALS, a type of bird, are changing over time. They might be VERY different millions of years in the future. | 4.23 (0.93) | 4 | 1-5 |
| Most living things today are VERY different from their ancestors who lived a long, long time ago. | 4.07 (1.03) | 4 | 1-5 |
| HUMAN BEINGS, a type of primate, are changing over time. They might be VERY different millions of years in the future. | 4.01 (1.07) | 4 | 1-5 |
| Evolution is still going on TODAY. | 4.66 (0.73) | 5 | 1-5 |
| COYOTES, a type of mammal, are changing over time. They might be VERY different millions of years in the future. | 4.19 (0.97) | 4 | 1-5 |
| **Evolution Composite** | **4.23 (0.77)** | **4.4** | **1-5** |

Note: Ratings on a 5 point scale. N=125

Table 64: Self-reported ratings of exhibit experience, DeepTree Naturalistic Study

| Question | N | Mean (SD) | Median |
|---|---|---|---|
| How interesting was the touch table? | 33 | 4.24 (0.83) | 4 |
| How much did you like using the touch table? | 33 | 4.15 (0.94) | 4 |
| How much did you learn at the touch table? | 33 | 3.27 (0.94) | 3 |
| How did others at the table affect your learning? | 23 | 3.09 (0.95) | 3 |
| How much did you like working with others at the table? | 24 | 3.46 (0.98) | 3.5 |
| Would you tell a friend to visit the table? | 33 | 4.24 (1.17) | 5 |

Note: Ratings on a 5 point scale. N=33.

Table 65: Levels of Agreement with Survey Questions about Common Ancestry of Various Species. DeepTree Naturalistic Study

| Species Pairs | Mean (SD) | Median | Min- Max |
|---|---|---|---|
| Rabbits and Lizards | 3.77 (1.45) | 4 | 1-5 |
| Humans and Mushrooms | 3.63 (1.61) | 4.5 | 1-5 |
| Mice and Rats | 4.43 (1.10) | 5 | 1-5 |
| Bears and Sunflowers | 3.37 (1.73) | 4 | 1-5 |
| **Common Ancestor Composite** | **3.80 (1.21)** | **3.9** | **1-5** |

Note: Ratings on a 5 point scale. N=30

Table 66: Levels of Agreement with Survey Questions about Evolution. DeepTree Naturalistic Study

| Statement | Mean (SD) | Median | Min- Max |
|---|---|---|---|
| CARDINALS, a type of bird, are changing over time. They might be VERY different millions of years in the future. | 4.16 (1.00) | 4 | 1-5 |
| Most living things today are VERY different from their ancestors who lived a long, long time ago. | 4.19 (1.22) | 5 | 1-5 |
| HUMAN BEINGS, a type of primate, are changing over time. They might be VERY different millions of years in the future. | 4.29 (0.86) | 4 | 1-5 |
| Evolution is still going on TODAY. | 4.61 (0.67) | 5 | 3-5 |
| COYOTES, a type of mammal, are changing over time. They might be VERY different millions of years in the future. | 4.13 (0.99) | 4 | 2-5 |
| **Evolution Composite** | **4.28 (0.78)** | **4.2** | **1.6-5** |

Note: Ratings on a 5 point scale. N=31

## *BAT Survey Results*

Table 67: Self-reported ratings of exhibit experience, BAT Video Study

| Question | N | Mean (SD) | Median |
|---|---|---|---|
| How interesting was the touch table? | 18 | 4.28 (0.46) | 4 |
| How much did you like using the touch table? | 18 | 4.44 (0.62) | 4.5 |
| How much did you learn at the touch table? | 19 | 3.89 (0.94) | 4 |
| How did others at the table affect your learning? | 14 | 4.29 (0.83) | 4 |
| How much did you like working with others at the table? | 14 | 4.50 (0.52) | 4.5 |
| Would you tell a friend to visit the table? | 19 | 4.16 (1.07) | 4 |

Note: Ratings on a 5 point scale. N=19.

Table 68: Levels of Agreement with Survey Questions about Common Ancestry of Various Species. BAT Video Study

| Species Pairs | Mean (SD) | Median | Min- Max |
|---|---|---|---|
| Rabbits and Lizards | 3.89 (1.24) | 4 | 1-5 |
| Humans and Mushrooms | 3.11 (1.78) | 4 | 1-5 |
| Mice and Rats | 4.63 (0.60) | 5 | 3-5 |
| Bears and Sunflowers | 2.89 (1.66) | 3 | 1-5 |
| **Common Ancestor Composite** | **3.65 (1.05)** | **3.75** | **2-5** |

Note: Ratings on a 5 point scale. N=19.

Table 69: Levels of Agreement with Survey Questions about Evolution. BAT Video Study

| Statement | Mean (SD) | Median | Min-Max |
|---|---|---|---|
| CARDINALS, a type of bird, are changing over time. They might be VERY different millions of years in the future. | 4.68 (0.48) | 5 | 4-5 |
| Most living things today are VERY different from their ancestors who lived a long, long time ago. | 4.32 (1.00) | 5 | 2-5 |
| HUMAN BEINGS, a type of primate, are changing over time. They might be VERY different millions of years in the future. | 4.42 (0.96) | 5 | 2-5 |
| Evolution is still going on TODAY. | 4.63 (0.96) | 5 | 1-5 |
| COYOTES, a type of mammal, are changing over time. They might be VERY different millions of years in the future. | 4.63 (0.50) | 5 | 4-5 |
| **Evolution Composite** | **4.54 (0.51)** | **4.8** | **3.6-5** |

Note: Ratings on a 5 point scale. N=19.

Table 70: Self-reported ratings of exhibit experience, BAT Naturalistic Study

| Question | N | Mean (SD) | Median |
|---|---|---|---|
| How interesting was the touch table? | 12 | 4.33 (0.65) | 4 |
| How much did you like using the touch table? | 12 | 4.50 (0.67) | 5 |
| How much did you learn at the touch table? | 12 | 3.67 (0.65) | 4 |
| How did others at the table affect your learning? | 9 | 3.67 (1.12) | 3 |
| How much did you like working with others at the table? | 10 | 4.20 (1.23) | 4.5 |
| Would you tell a friend to visit the table? | 12 | 4.25 (0.87) | 4.5 |

Note: Ratings on a 5 point scale. N=12.

Table 71: Levels of Agreement with Survey Questions about Common Ancestry of Various Species. BAT Naturalistic Study

| Species Pairs | Mean (SD) | Median | Min- Max |
|---|---|---|---|
| Rabbits and Lizards | 3.09 (1.22) | 3 | 1-5 |
| Humans and Mushrooms | 2.42 (1.44) | 2.5 | 1-5 |
| Mice and Rats | 4.50 (0.90) | 5 | 2-5 |
| Bears and Sunflowers | 2.42 (1.44) | 2.5 | 1-5 |
| **Common Ancestor Composite** | **3.09 (0.91)** | **3** | **2-5** |

Note: Ratings on a 5 point scale. N=12.

Table 72: Levels of Agreement with Survey Questions about Evolution. BAT Naturalistic Study

| Statement | Mean (SD) | Median | Min- Max |
|---|---|---|---|
| CARDINALS, a type of bird, are changing over time. They might be VERY different millions of years in the future. | 3.33 (1.30) | 4 | 1-5 |
| Most living things today are VERY different from their ancestors who lived a long, long time ago. | 3.75 (1.60) | 4.5 | 1-5 |
| HUMAN BEINGS, a type of primate, are changing over time. They might be VERY different millions of years in the future. | 3.83 (1.40) | 4 | 1-5 |
| Evolution is still going on TODAY. | 4.42 (1.16) | 5 | 1-5 |
| COYOTES, a type of mammal, are changing over time. They might be VERY different millions of years in the future. | 3.83 (1.27) | 4 | 1-5 |
| **Evolution Composite** | **3.83 (1.18)** | **4.2** | **1-5** |

Note: Ratings on a 5 point scale. N=12.