

Insights From Experiments With Rigor in an EvoBio Design Study

Jen Rogers, Austin H. Patton, Luke Harmon, Alexander Lex, Miriah Meyer

Abstract— Design study is an established approach of conducting problem-driven visualization research. The academic visualization community has produced a large body of work for reporting on design studies, informed by a handful of theoretical frameworks, and applied to a broad range of application areas. The result is an abundance of reported insights into visualization design, with an emphasis on novel visualization techniques and systems as the primary contribution of these studies. In recent work we proposed a new, interpretivist perspective on design study and six companion criteria for rigor that highlight the opportunities for researchers to contribute knowledge that extends beyond visualization idioms and software. In this work we conducted a year-long collaboration with evolutionary biologists to develop an interactive tool for visual exploration of multivariate datasets and phylogenetic trees. During this design study we experimented with methods to support three of the rigor criteria: ABUNDANT, REFLEXIVE, and TRANSPARENT. As a result we contribute two novel visualization techniques for the analysis of multivariate phylogenetic datasets, three methodological recommendations for conducting design studies drawn from reflections over our process of experimentation, and two writing devices for reporting interpretivist design study. We offer this work as an example for implementing the rigor criteria to produce a diverse range of knowledge contributions.

Index Terms—Methodologies, Application Motivated Visualization, Guidelines, Life Sciences Visualization, Health, Medicine, Biology, Bioinformatics, Genomics

1 INTRODUCTION

Design study is an established approach to problem-driven visualization inquiry that emphasizes designing visual analysis tools in close collaboration with domain experts [66]. Within a design study, visualization researchers build an understanding of a problem domain and translate that understanding into a visualization design, iteratively refining both their understanding of the problem and their visual analysis solution through close work with domain collaborators. Researchers conducting design studies draw from a host of theoretical constructs to guide the inquiry process, from process models [23, 39, 42, 44, 66] to design decision models [46, 50], guiding scenarios [65], educational models [72], and collaboration roles [69, 78]. As a result, an increasing number of reports describe effective design studies within a broad range of application areas [9, 26, 32, 40, 41, 52, 56, 82].

Historically, design study papers have emphasized novel visual analysis systems and techniques as primary knowledge contributions [44]. Many of these papers also cite domain characterizations and abstractions [50] as contributions under the reasoning that they are important for judging the validity of technical design artifacts and for building a body of visual analysis requirements that others can design against. The original definition of design study also includes lessons-learned as a potential knowledge contribution stemming from reflection, but scant guidance is available on how to generate knowledge of this sort [43].

In Meyer & Dykes [44] we proposed a new, interpretivist view of visualization design study to produce a more diverse range of knowledge contributions. As a critique of the software-centric view of design study, this new perspective emphasizes the potential for using design study to acquire a more diverse range of knowledge, including knowledge about the visualization design process as well as about people’s relationship with data and technology more broadly. This work recommends six rigor criteria for guiding the design study process toward acquiring new knowledge: INFORMED, REFLEXIVE, ABUNDANT, PLAUSIBLE, RESONANT, and TRANSPARENT. These criteria provide an opportunity for

researchers to rethink how to conduct effective design studies, learning new things along the way.

In the work we present here we experimented with methods to support three of the rigor criteria: ABUNDANT, REFLEXIVE, and TRANSPARENT. Our experimentations took place within the context of a one-year design study with evolutionary biologists. We employed techniques such as an immersive, three-month field study; structured and systematic reflection; and careful curation of documents and other design artifacts. Through a period of collaborative, critical reflection, we identified several methodological insights that emerged from our experiments.

The resulting contributions from this inquiry are diverse, including both technical and methodological insights. More specifically, the contributions include:

- Two new visualization techniques for supporting the analysis of multivariate trees: (1) a *trait view* that visualizes node-value distributions under uncertainty for associated characteristics along multivariate subtrees; and (2) a *pattern view* that aids in the discovery and visualization of patterns in value trajectories for attributes across paths in a tree.
- Three methodological recommendations for conducting interpretivist design study: (1) establish systematic reflective practices that include reflexive notes, reflective transcriptions, and artifact curation; (2) build and maintain a trace of diverse research artifacts; and (3) argue for rigor from evidence, not just methods.
- Two experimental writing devices for reporting on interpretivist design study: (1) inclusion of direct links to research artifacts to transparently provide an abundance of evidence; and (2) embedding of a design study paper within a methodological one to highlight the diversity of our research contributions.

This work serves as an example of how researchers can consider the ABUNDANT, REFLEXIVE, and TRANSPARENT criteria in practice, as well as the diverse types of knowledge contributions possible through their consideration.

We first provide the theoretical backdrop for our methodological work in Section 2, followed by a description of our research methods in Section 3. Section 4 is a design study paper-within-a-paper, emphasizing the technical aspect of this work; our methodological recommendations follow in Section 5. Throughout the paper we include direct links to our abundant collection of research artifacts — for example [T45] — to transparently provide evidence for our claims.

2 THEORETICAL BACKDROP

The methodological work we present in this paper draws from the interpretivist perspective of design study proposed by Meyer & Dykes [44].

• Jen Rogers, Alexander Lex, Miriah Meyer are with the University of Utah, E-mail: jen@sci.utah.edu, alex@sci.utah.edu, miriah@cs.utah.edu.

• Austin H. Patton is with Washington State University, E-mail: austin.patton@wsu.edu.

• Luke Harmon is with the University of Idaho, E-mail: lukeh@uidaho.edu.

This is the authors preprint version of this paper. Please cite the following reference: Jen Rogers, Austin H. Patton, Luke Harmon, Alexander Lex, Miriah Meyer. Insights From Experiments With Rigor in an EvoBio Design Study. IEEE Transactions on Visualization and Computer Graphics, to appear, 2020.

This perspective argues for a myriad of opportunities for researchers to make valuable knowledge contributions beyond visualization techniques and software. Doing so, however, requires a rethinking of design study research practices and the ways we make quality judgments about the inquiry. Six criteria for rigor guide an interpretivist design study approach — INFORMED, REFLEXIVE, ABUNDANT, PLAUSIBLE, RESONANT, and TRANSPARENT — which are derived from theoretical positions in social science [35, 70, 76], information systems [67], and research through design [20, 84]. Achieving all six criteria within a single design study is unlikely to occur due to pragmatic constraints such as time and resources [44]. In the work presented in this paper, we focus on ABUNDANT, REFLEXIVE, and TRANSPARENT, exploring various ways to achieve these criteria, as well as the kinds of knowledge elucidated by doing so.

2.1 Abundant

A design study with abundance reflects the richness and complexity of the situation under study [44]. An abundant design study thus includes a rich and diverse body of evidence, as well as an abundance of other considerations such as participant voices, designs, and time in the field. In our experiments we considered all of these aspects of abundance.

The inclusion of a variety of voices and contexts reflects a valuing of **pluralism** found in critical feminist theory that “insists that the most complete knowledge comes from synthesizing multiple perspectives” [30]. In human computer interaction (HCI), pluralism is argued as a mechanism for resisting designs that embed “any single, totalizing, or universal point of view” [3]. Arguments for pluralism can be grounded in the idea of situated knowledges [24], which argues an epistemic view of a singular reality that can only be known only partially, embedded within a specific context. It is by combining these partial perspectives — through “actively and deliberately inviting other perspectives into the data analysis” [30] — that a researcher achieves a fuller, richer view of the situation under study.

An emphasis on exploring a design space through many, rapid designs similarly helps a designer avoid blind spots and fixation on a singular solution [12, 15]. Design problems are wicked by nature, with an extensive space of possible solutions [11]. By broadly considering a design space, designers are more likely to find good solutions, rather than average or poor ones [66], as well as to develop a better understanding of the problem under study [15]. Dow et al. recommend exploring and refining design ideas in parallel, rather than through a sequential process, to obtain better and more diverse design artifacts [18]. In the same vein, Buxton advocates for rapid sketching with broad ideation for developing effective design concepts through iterations of “controlled convergence” [12].

Finally, abundance through prolonged engagement with the people and context under study is a mainstay of qualitative research [35, 68, 76]. Researchers who establish an early familiarity with a domain build trust with their participants as well as the ability to understand domain-specific nuances of what they observe: “objects and behaviors take not only their meaning but their very existence from their contexts” [35]. In a visualization study, *design by immersion* is an approach for engagement in which both the visualization researchers and domain experts “participate in the work of another domain such that visualization design, solutions, and knowledge emerge from these transdisciplinary experiences and interactions” [23]. This methodology allows visualization researchers to enrich their understanding of a domain, explore a broader visualization solution space, and build trust and agency with collaborators. Field studies — in which a researcher spends sustained time with participants in their natural environment — is a technique that can support visualization researchers in achieving immersion through prolonged engagement [40].

2.2 Reflexive

Being reflexive within a visualization design study is to strive for “explicit and thoughtful self-awareness of a researcher’s own role in a study” [44]. As a cornerstone of interpretivist, qualitative research, reflexivity is an acknowledgement of a researcher’s influence on a study, and vice versa [4]. Researcher bias and perspective are an inherent part of qualitative research, and eliminating them from the research process

is arguably impossible [38]. Reflexivity is instead an opportunity to gather valuable data [61] that can help researchers understand their biases and perspectives as a vector for change and learning [19].

Reflexivity is an important consideration in the third wave of HCI research [6]. Largely discussed in the critical HCI literature, reflexivity is considered a mechanism for researchers “to be accountable for the ways in which HCI construes design(ing) and acknowledge our responsibility … to challenge the dominant view on design” [2]. Despite its importance, the HCI community has been slow to broadly adopt reflexive practices in research due to the scrutiny on subjectivity during the review process. **The visualization research community shares a similar emphasis and valuing of objectivity [44], and a lack of methods for supporting and exploiting reflexivity. This gap motivated our experimentations with reflexivity.**

Reflexivity is a type of (self) reflection [37]. As a method, reflection traces to Schön’s ideas of reflective practice through reflection-in-action and reflection-on-action [64]. Reflection-in-action is characterized as an intuitive, rapid, reflective response “in the moment” [80]. Reflection-on-action instead happens after an experience, and is characterized as an “inquiry into the personal theories that lie as the basis of one’s actions” [31]. A commonly employed method for reflection-on-action in qualitative research is *memoing*: “Memos can help to clarify thinking on a research topic, provide a mechanism for the articulation of assumptions and subjective perspectives about the area of research, and facilitate the development of the study design” [4]. We used memoing throughout our design study to facilitate reflexivity and reflection.

Pragmatically, reflection-on-action is synonymous with critical reflection [16], an inquiry process where researchers question their assumptions by examining the reasoning and ideology that frame their practice and experiences [10, 75]. Work by Kerzner et al. employs critical reflection to construct a general framework for visualization workshops from their experiences running 17 of them [29]. Similarly, Satyanarayan et al. create a set of lessons for designing visualization authoring toolkits using what they call critical reflections [63]. Although not grounded in the reflection literature, their process is similar to that of reflection-on-action practices. Other than a handful of examples like these, the visualization literature is largely lacking pragmatic guidance on how and when to reflect [43]; **this work contributes actionable recommendations for reflecting in a design study.**

2.3 Transparent

Transparent reporting of a design study — through **scrutinizable documentation of data, methods, analysis, and artifacts** — is necessary for supporting judgments about the quality of the study and its results [44]. How to report transparently, however, is an open question. Recent work by Wacharamantham et al. provides recommendations for sharing HCI research materials based on a survey of researchers [77]. This work, however, considers only software and hardware prototypes for design-oriented studies, missing many of the diverse artifacts produced within a design study such as sketches, abstractions, reflexive notes, and diagrams. In this work we experimented with recording and reporting a diverse set of design artifacts, drawing from ideas in qualitative research and research through design.

In interpretivist, qualitative research, the *audit trail* is an established mechanism for transparent reporting [1, 13, 17, 35]. An audit trail is a detailed documentation of a research process that is intended for use in an *audit process* [1]. This process is undertaken by an (external) auditor who reviews the audit trail in order to assess the quality of the study, enhancing the trustworthiness of the research [35]. Although audit trails are meant to increase the transparency of a study, they can also increase the quality through explicit thoughtfulness on the part of the researcher on what and how to record [17]. **Two recent visualization design studies include audit trails as supplemental materials [29, 40], but neither study performed an audit.**

Transparently reporting on design decisions and insights is challenging due to the ingrained nature of knowledge within the artifacts themselves. Design scholars consider the knowledge that a designer acquires to reside in the artifacts they create [14]. This knowledge, however, is implicit and often opaque [71]. *Annotated portfolios* — textual

annotations of design patterns across a curated collection of designs — is a method used within the research-through-design community to explicitly communicate knowledge embedded within designs [8, 21]. Annotations allow for comparison of designs and highlight relationships between disparate works, from which designers can develop and communicate generalized, intermediate knowledge. A different approach to externalizing design knowledge is that of literate visualization, which engages the designer in reflective documentation during the creation of digital, visualization artifacts [79].

3 METHODS

To explore how an interpretivist approach to design study changes what and how we learn, we set out with the goal of experimenting with three criteria — ABUNDANT, REFLEXIVE, and TRANSPARENT — during an evolutionary biology design study. We positioned this work within the perspective that design studies are wicked, subjective, and diverse [44]. Rogers conducted a three-month, immersive field study, followed by a design phase and a reflection phase in collaboration with Lex and Meyer. In this section we provide details about our research site and domain collaborators, the ways we experimented with the criteria, and the methods we employed for data collection and analysis. We directly link to our abundant collection of evidence — for example [T45] — to provide transparent reporting of our process.

3.1 Research Site and Participants

Our study took place at two sites. In the first phase, we undertook a three-month field study in the Harmon Lab at the University of Idaho, which studies ecology and evolution through phylogenetic analysis. During this time, Rogers spent work-hours within the group’s lab, immersed in conditions similar to those in which the evolutionary biology graduate students worked. The lab environment was open and social, with six desks spaced around the edges of the room, a community couch often inhabited by other graduate students who stopped by, and a white board filled with scattered drawings and notes. The graduate students used this space for their computational work, which was often analysis of the phylogenetic data and field sample measurements taken from summer field work. This lab was chosen based on a relationship established through a federally funded research project [45] between the Harmon Lab and the Visualization Design Lab at the University of Utah. The design and reflection phases took place within the Visualization Design Lab.

During the field study we worked with seven evolutionary biology collaborators. Two primary collaborators during this phase were Harmon, the PI of the evolutionary biology lab, and Patton, a graduate student at Washington State University who works closely with the Harmon lab, often on-site. Both primary collaborators are co-authors on this paper. Five other graduate students in the lab served as secondary collaborators. All collaborators were involved with the interviews and informal feedback. The primary collaborators were additionally involved with the design and evaluation of our visualization techniques.

3.2 Criteria Considerations

Our decision to focus on the ABUNDANT, REFLEXIVE, and TRANSPARENT criteria stemmed from our experiences in previous studies and considerations of actionability [T160]. In previous work we attempted to instill transparency through collecting artifacts and releasing audit trails [29, 40]. These experiences led to numerous conversations within our research group about how to record and report artifacts in design studies and other qualitative research studies. We saw this design study as an opportunity to systematically experiment with abundant recording and transparent reporting of evidence from the very start of a study. We included reflexivity based on the interests of the research team and the actionability of reflexive memoing. Our approaches to meeting these criteria evolved over the course of the study.

We attempted to instill abundance in our design study in four ways. First, we meticulously curated a rich collection of artifacts generated throughout the design study including field notes and reflective memos [T48], email correspondence [T90], sketchbook scans [T81], photos

of collaborator sketches [T55], links to papers [T87], low- and high-fidelity visualization prototypes [T158, T96], and notes reflectively transcribed from audio recordings of meetings [T36]. Second, we conducted an immersive field study, in which Rogers situated herself as a peer in the Harmon Lab for three months. Working in the communal space of our domain collaborators, Rogers actively engaged in research meetings and reading clubs focused on evolutionary topics of interest. She learned how to use the analysis pipelines of her collaborators to get a deeper understanding of the domain problem space [T47, T50]. Through time, she gained a deeper understanding of the domain research and developed a personal investment in our collaborators’ research and social dynamics. These activities encompass the *communal, personal, and active* themes of immersive studies [23]. Third, we contacted domain experts outside the Harmon Lab in an attempt to include multiple voices and datasets. We sent emails to colleagues of the Harmon Lab, as well as evolutionary biology researchers at the University of Utah, inviting them to participate in the evaluation of our visualization designs [T109]. Fourth, we relied heavily on sketching to facilitate brainstorming of visualization ideas [T43, T52], to understand the domain space [T10, T38], to communicate with domain collaborators [T55], and to aid in reflective analysis [T138].

We implemented reflexivity during the field study through regular, reflective memoing by Rogers. These reflections were reflexive in nature and included documenting her feelings as she became more incorporated into the lab, her insecurities that were potentially limiting the research [T3, T20], her interpretations on social dynamics and friendships within the lab, and how those dynamics affected the research [T18]. Memoing was done before and after meetings and during pivot-point moments in the research process.

In an attempt to transparently communicate the design study process, we created an auditable website from our collection of research artifacts, which is available at <http://vdl.sci.utah.edu/trrrace/>. This website which we call a *trrrace* and discuss in more detail in Section 5.2, traces the project from the field study through the design and reflection phases, organizing the abundant collection of artifacts we recorded throughout. The artifacts are organized in an interactive timeline and are discoverable via annotations, descriptive metadata, and directly in the timeline.

3.3 Data Collection

We kept a meticulous collection of all recorded artifacts starting from the beginning of the field study in an attempt to record an abundance of evidence from our design study process and support transparency. These artifacts were generated throughout all three phases of research, but the content creation was concentrated during times of immersion in the field study, as well as during times of correspondence with collaborators in the design phase of the tool. Throughout the field study, Rogers interviewed members of the lab, taking reflective notes before and after every interview. Pre-interview reflections included a review of previous meeting notes and outlining an agenda [T8], and post-interview reflections summarized the main talking points and speculated about productive next steps [T20]. Additionally, she audio-recorded these interviews and reflectively transcribed [40] them to capture the context of what was said when, how things were said, and her interpretation of the conversations [T53]. To capture a rich view of the interviews, Rogers recorded any white-board diagrams [T94], scribbles [T41], or sketches [T55] that were generated during discussions. In addition to the pre- and post-interview reflections, Rogers also regularly wrote reflexive memos that included her feelings on her immersion in the lab, her insecurities that were possibly limiting the research, friendships, social dynamics, and how those dynamics affected the research [T3, T18, T20].

During the second week of the field study, Rogers conducted a creative visualization opportunities workshop [29] with the lab members to brainstorm about potential visualization directions. We took photos of all the materials generated from the workshop exercises and audio recorded the workshop [T23, T24, T25, T26, T27, T28, T29, T30, T31, T32, T33, T34, T35, T36].

The beginning stages of sketching and prototyping began during

the field study, but the bulk of the design work and tool development happened during the design phase. Our primary collaborators remained extensively involved in providing feedback on design iterations, with much of this feedback happening through video calls, email, and in two, short, subsequent visits to the Harmon Lab. We recorded feedback emails [T90, T118], notes from the in-person feedback sessions [T125], and memos capturing personal interpretations of the feedback [T126]. Design artifacts generated during this process include sketches [T43, T45, T52], mock-ups [T59], and screen-shots of prototype iterations [T67, T73, T92].

3.4 Analysis

Analysis occurred during the final, reflective stage of the study when we started the construction of an audit trail as a website for collecting and annotating our diverse set of research artifacts. The website was initially designed to communicate the design study process with a high-level of transparency and detail. The organization and curation of artifacts, however, became a powerful catalyst for reflection that led to significant methodological insights about our design process, as well as new directions for the design of the visualization tool. Through collaborative, critical reflection among the visualization research team members, we iteratively developed a set of actionable recommendations for conducting interpretivist design study from our insights looking across the collection of artifacts.

4 TREVO: AN EVOLUTIONARY BIOLOGY DESIGN STUDY

This design study was motivated by the complexity of our collaborators' problem in representing the rich, multivariate, and uncertain data in their analysis. They work extensively with trees that represent hypothesized explanations for how species are related. In this design study we developed a web-based visualization tool Trevo, that allows them to analyze these trees with multivariate and uncertain attributes.

We report on this design study in an abbreviated form as a paper-within-a-paper as part of our larger goal of highlighting the diverse contributions possible from interpretivist design study. This experimental format emerged from our dissatisfaction with early paper drafts that followed a more traditional design study reporting structure [T144, T159]. We felt the traditional structure overly accentuated technical contributions while leaving little room for significant methodological discussions. We developed the paper-within-a-paper style to stress the role of the design study as a method of inquiry [44] that reflects and reports on a more diverse type of knowledge.

4.1 Biological Background

The driving question in the field of evolutionary biology is why the living world evolved the way it did? To answer this question, researchers need to determine when a given trait evolved, such as a lizard's long tail, and whether a particular species possesses that trait as a result of common ancestry or of other forces such as the environment. To answer these questions, evolutionary biologists study a group of living organisms to establish hypotheses about evolutionary forces that can generalize to other species. For example, researchers study anole lizards to infer how environment influences evolution. Analysis begins in the field, where these researchers take samples of living species and measure their physical characteristics, such as a lizard's tail length, snout length, and body mass. They use these measurements of current species, typically along with DNA sequence data, to reconstruct physical characteristics of the ancestors in a species' phylogenetic history. These histories are then the basis of studying when and why traits evolved, and whether the physical characteristics of contemporary species are, or are not, a result of evolution from common ancestors.

Evolutionary relationships are commonly represented as a binary tree, referred to as a **phylogenetic tree**. These trees are usually reconstructed by modeling the evolution of a set of DNA sequences sampled from present-day species. The leaf nodes of the tree represent the contemporary species, whereas inner nodes represent their common ancestors. All nodes in the tree have associated characteristics described by a set of traits. Internal nodes (common ancestors) have estimated

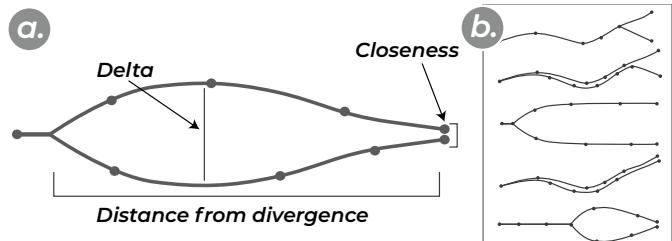


Fig. 1. Defined preset patterns in the pattern view. (a) Pattern breakdown for convergence (b) The six predefined patterns.

values for these traits. Leaf nodes (species) have measured values for traits.

A common structure evolutionary biologists work with is **clades**, which are subtrees of the larger phylogeny in which all species share a single, unique, common ancestor. For example, for anole lizards, the main clade of study is the genus *Anolis*, a group of more than 400 species that all evolved from a common ancestral lizard. These subtrees are sometimes predefined, as is the case for well-established clades such as anoles, or they can be defined during analysis.

Researchers analyze different, possible evolutionary mechanisms by studying **patterns of evolution**. These patterns can be summarized in terms of how traits change, or evolve, along the branches of the phylogeny. A common pattern of trait evolution is that of **divergence** in which species evolve increasingly distinct trait values over time [60]. Another pattern, **convergence** which is shown in Figure 1(a), is characterized by traits that diverge early in two species' histories, but then converge later in their evolutionary histories by developing similar or identical traits [28]. Convergence is an indication of adaption — certain traits evolve repeatedly because they are beneficial in an environment — and has been studied extensively in the anole lizards. Many of these lizards, having split off from their common ancestors a long time ago, inhabit similar environments on separated islands and have evolved very similar characteristics as a consequence. Although other interesting patterns besides divergence and convergence exist, such as those in Figure 1(b), they do not have standardized names.

Identifying patterns of evolution is a challenging analysis problem that involves accounting for changes to multiple traits under uncertainty in the context of the tree topology. We worked with our collaborators to explore new ways to enable this complex analysis with interactive visual analysis tools.

4.2 Data and Task Abstraction

In the datasets our collaborators are analyzing, evolutionary relationships are represented as rooted trees. Bifurcations in the tree represent speciation events. Internal nodes encode hypothesized common ancestors of existing species, which in turn constitute the leaf nodes. The size of the trees we focused on here ranged from 20 to 200 species (leaves), each associated with 5 to 25 traits. Traits of a species can be discrete or continuous and are uncertain for the reconstructed (inner node) species. Reconstructed discrete traits, such as the geographic location where a species is found or whether they lay eggs, are specified as probabilities. Continuous traits, such as tail length, are given as an estimated value and a 95% confidence interval.

To explain why the living world evolved the way it did, our collaborators' analysis is focused on understanding when and how traits evolved in a population, which requires viewing trait values for multiple attributes in the context of the topology of the tree. We break down this larger analysis goal into three domain tasks:

T1: Understand the uncertainty in multiple reconstructed traits.

Significant uncertainty exists in the reconstructed traits for internal nodes, so adequate visual representations of trait values and their uncertainty are critical. Current methods for visualizing attributes in phylogenetic trees are limited to showing one or two traits at a time, and frequently cannot encode uncertainty [T42, T36, T16]. This task is orthogonal to all other tasks, i.e., uncertainty analysis is a part of every analysis task.

T2: Analyze subtrees.

This task is concerned with creating and analyzing individual subtrees (clades) and comparing between multiple subtrees.

T2.1: Create subtrees.

Our collaborators need the ability to create subtrees by topology and trait values. For example, an analyst might want to create two subtrees based on an attribute, such as the island a species is inhabiting [T64, T80]. Definitions of subtrees might also be given as formal clades in a dataset.

T2.2: Analyze attribute distributions in subtrees.

Our collaborators need to be able to identify significant changes in multiple traits at once. For example, understanding whether a shift toward a longer tail is correlated with a shift toward longer hind-legs can give hints about the underlying causes of that change [T20]. Viewing multiple traits at once is particularly difficult for our collaborators, who rely on comparisons of reconstructed traits on separate trees [T36, T72].

T2.3: Identify evolutionary outliers.

It is important for our collaborators to identify individual species, paths, or subtrees that have significantly different trait values compared to the rest of the subtree [T17, T91]. For example, they want to identify paths with species that have a larger body mass than the rest of the subtree.

T2.4: Compare attribute distributions of multiple subtrees.

Comparisons are important in characterizing what makes a subtree unique. For example, our collaborators want to study whether the species in a subtree share common characteristics, such as head and tail length, that set them apart from the rest of the tree. To study how traits evolved through history, they need to understand how subtree trait distributions diverge and where this happens in the tree [T4, T20, T66, T72, T88].

T3: Identify and analyze evolutionary patterns

An important task in our collaborators' analysis is identifying the evolutionary patterns that indicate certain mechanisms underlying evolution [T53, T64, T87]. Identifying these patterns requires the comparison of trait trajectories of multiple species in a tree [T80, T93]. To identify convergence, for example, an analyst would search for two paths that separated early in the tree with trait values that first diverged, but then later converged.

4.3 Related Work

Visualization of phylogenetic data is challenging in three ways: (1) the trees can be large, requiring sophisticated navigation and/or aggregation strategies to browse them; (2) the topology of the trees is uncertain, requiring the comparison of multiple alternative trees; and (3) the trees are associated with many (uncertain) attributes, requiring sophisticated multivariate tree visualization strategies. Our work addresses the third problem, multivariate trees, but we briefly review all areas.

The scale and uncertainty of topology remain challenges in phylogenetic research and numerous visual solutions have been proposed for both [5, 7, 33, 34, 36, 51, 62]. Large phylogenetic trees and topological uncertainty are not key problems for our collaborators; visualizing trees with many attributes, however, is. As a generalization, visualizing many traits in the context of a tree is a type of multivariate network visualization problem. Nobre et al. recently described the design space of a multivariate network visualization in a survey that included tree visualization [53]. We here focus mostly on approaches for phylogenies but refer readers to this survey for a broader overview.

Within the evolutionary biology community, visualizations of phylogenetic data are used for both exploration and presentation in papers. Most figures found in evolutionary biology papers show trees laid out using node-link diagrams with either linear or circular layouts, and on-node or on-edge encoding to show trait values [58, 60]. These figures are often created with interactive tools such as iTOL [34] or Dendroscope [27], or using scripted plotting libraries, such as phytools or ggtree for R [57, 81]. Tools such as iTOL can visualize multiple attributes for the leaves, but the inner nodes are usually limited to a single attribute. Analysts, however, often need to account for multiple traits at once to identify underlying forces influencing trait change. In their current workflow, they compare different traits mapped to the

nodes of multiple trees side-by-side. Such comparisons are difficult with just 2 traits, but analysts must often consider up to 10 traits for a given tree. As expressed by one of our collaborators, "if you have 1 continuous trait you can do things. If you have 2 — OK. If you have 3 or 4 or 5, there is nothing really sufficient" [T36].

In the visualization community, several tools have been designed to visualize trees with attributes. Lineage [52], for example, visualizes attributes for genealogical trees using a linearization approach, where the attributes are shown in a table; Juniper is a generalization of this method to networks [54]. Other tools, such as TreeVerty2 [22], visualize attributes using implicit layouts. Researchers currently have no tool suitable for visualizing many traits for inner nodes and leaves under uncertainty in the context of phylogenetic trees.

4.4 Visualization Design

Two technical contributions emerged from this design study. The first is a technique for visualizing summary distributions of attributes in a (sub)tree — the **trait view** — designed to address the analysis of subtrees (T2). The second contribution is a view for querying, ranking, and visualizing patterns consisting of topological and attribute features — the **pattern view** — designed to address the identification and analysis of evolutionary patterns (T3). Both views visualize uncertainty (T1) and were implemented in a web-based tool we call Trevo, along with two additional views: <https://vdl.sci.utah.edu/Trevo/>.

4.4.1 Trait View

A crucial task for our collaborators is analyzing patterns of attributes within and between subtrees. When subtrees are defined topologically, this analysis can be supported in the context of a phylogenetic tree. For subtrees defined based on trait values, however, species can be scattered across a phylogenetic tree. For example, our collaborators want to create two subtrees for anole species that are found on the islands of Hispaniola and Cuba so they can compare the distribution of body mass of the lizards on these islands to study any environmental effects that might appear. The "island" trait does not clearly split the phylogenetic tree into disjunct subtrees, as common ancestors colonized islands multiple times. It instead creates trees with partially overlapping branches. Figure-2(b) shows these disjunct subtrees with the species color coded by island. Lizards originating from Hispaniola are colored green, and those originating from Cuba are colored blue. Our collaborators compare the subtrees' trait values through the evolutionary history to determine when and how these groups began to diverge, for example, to determine if there is a difference in body mass between the two islands and when this divergence in traits began to occur along the evolutionary history. Identifying differences in value trends and when they occur within the phylogenetic tree can be difficult given the overlapping topology.

Through an iterative design process with our primary collaborators [T68, T74, T108, T114], we tackled this challenge with an aggregation solution for creating trait-defined subtrees. The key aspect of this new trait view is that it enables analysts to filter branches of the tree based on traits of the leaves. Figure 3 shows the steps involved in transforming a node-link tree layout into the trait view. Initially, the tree is filtered to include only extant species with a certain attribute such as the green leaves in Figure 3(a). We then leverage temporal information to bin the other nodes in the subtree by time, shown in Figure 3(b). The leaves are assigned a separate bin for which the uncertain discrete- and numerical-trait distributions are visualized in columns. Nodes are shown at the top of the bin; their horizontal position is driven by their time attribute, allowing analysts to compare multiple uncertain trait distributions in a temporal context unhindered by the tree's topology. Next, we use different encodings for leaf nodes with known trait values versus inner nodes with uncertain ones, shown in Figure 3(c). The known attributes of the leaves are encoded using histograms. For continuous uncertain traits we show the median plus a 95% confidence interval for the estimated values and a kernel density estimate plot. Finally, probabilities for uncertain discrete traits are represented in the trait view as separate one-dimensional dot plots for each state; to reduce



Fig. 2. Trait view showing four continuous and two discrete trait variables for 100 *Anolis* lizard species. (a) Outliers in the last SVL bin are brushed. A traditional phylogenetic tree view, shown on the right, can be used to define subtrees. (b) Leaf nodes can be color-coded by trait category. This detail view shows all leaf nodes color-coded by island of origin. These categories can be used to define subgroups by trait category or value, independent of the topology of the tree.

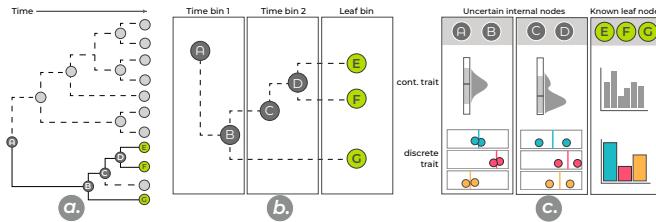


Fig. 3. Transforming a phylogenetic tree into the trait view. (a) We select a subtree by brushing for a trait in the leaves, shown in green. (b) The subtree is binned by time intervals and the leaves are assigned a separate bin. (c) We show continuous uncertain traits using a median plus confidence interval visualization and a KDE plot. For discrete uncertain traits we use multiple dot-plots, one for each trait category. Known traits are visualized using histograms.

the risk of overlapping dots, we use transparency and vertical jitter. The average for each state probability is plotted as a line in the plots.

4.4.2 Pattern View

The pattern view allows analysts to query for and find pairs of paths that follow a specific pattern of evolution such as convergence and divergence. Patterns of evolution are characterized by three key metrics: distance, delta, and closeness. The *distance* between two species refers to time and topological distance up to the first common ancestor. *Delta* is the maximum difference in an estimated continuous trait value after the species diverge. *Closeness* is the difference in a specific, continuous trait value between the extant species. We developed a query interface, shown in Figure 4(a), that analysts can use to define patterns of interest based on these three characterizing parameters. We found that while these simple parameters cannot represent arbitrary patterns, they covered all the patterns of evolution our collaborators are interested in. To simplify the pattern definition, we also developed six preset patterns that an analyst can choose from to score pairs of paths. These patterns, shown in Figure 1(b), emerged from repeated iterations with our collaborators [T94, T96, T129].

To create a ranking for paths that match a specified pattern we calculate scores for all possible pairs of leaves using the selected pattern parameters for all traits. We then rank the pairs of paths based on the initial trait chosen by the domain expert, and visualize the two paths using a ranked list of line+area charts, as shown in Figure 4(b). In this chart, the vertical axis corresponds to the trait value. Individual species are shown as squares, which are positioned to be centered on their most likely trait value. The height of the box shows the confidence interval. The boxes are connected by lines for the most likely value, and areas for the confidence interval.

One limitation of our original design of the pattern view was that

it could only show a single trait at a time [T96]. In an early feedback meeting, our collaborators asked if it was possible to have an indication of whether a specific pair of paths was also ranked highly for other traits [T99, T112, T113]. That is, in some cases the analysts might be interested in identifying species pairs that have converged in several traits, rather than just one. Convergence of sets of traits is of particular interest because such cases can provide the strongest evidence for adaptation to particular environments. To address this shortcoming, we added a supplementary heat map to the side of the pair plot that indicates whether the pair is ranked in the top 1% for a given pattern in any other traits in the data set, shown in Figure 4(b) on the right. Here, each square in a heat map represents other traits, where squares with darker saturation have a higher ranking. To find which pairs are ranked high for the pattern in the largest number of other traits, they can be sorted by frequency of top rankings from the heat map.

4.5 Case Study

We validate the trait and pattern views instantiated within Trevo by demonstrating their usefulness in a case study. The case study was conducted and written by our primary collaborators, who are also co-authors of this paper, and focuses on one of their primary datasets of the *Anolis* lizard genus. We provide a brief summary of findings here. We do not include the more detailed case study in this paper-within-a-paper, instead linking to it as external evidence [T145], as we find that domain-specific case studies often do not significantly contribute to a broader understanding of research contributions in design studies, but are rather akin to analysis scripts used in quantitative data analysis: they are necessary to ensure validity and trust, but do not convey knowledge on the subject of the research.

Using the trait view, our collaborators were able to reduce their analysis to a subset of species that exhibit exceptionally large body features, and to see how body features evolved differently over time. Traditional visualization approaches would have required coloring disjunct branches in a phylogenetic tree and making difficult judgments about color variations; the trait view instead provided targeted analysis using spatial encoding of the traits of interest. With the pattern view, our collaborators were able to confirm a known convergence and divergence event, a task not possible with commonly used software for the phylogenetic analysis of trait evolution. Furthermore, they were able to identify a new pattern of convergence in a pair of species, leading them to new biological questions about the evolutionary forces at play. This case study shows that our collaborators not only could easily distinguish interesting patterns in their data using Trevo, but also document a previously unknown insight. We offer this case study as evidence of the validity of our proposed designs [50].

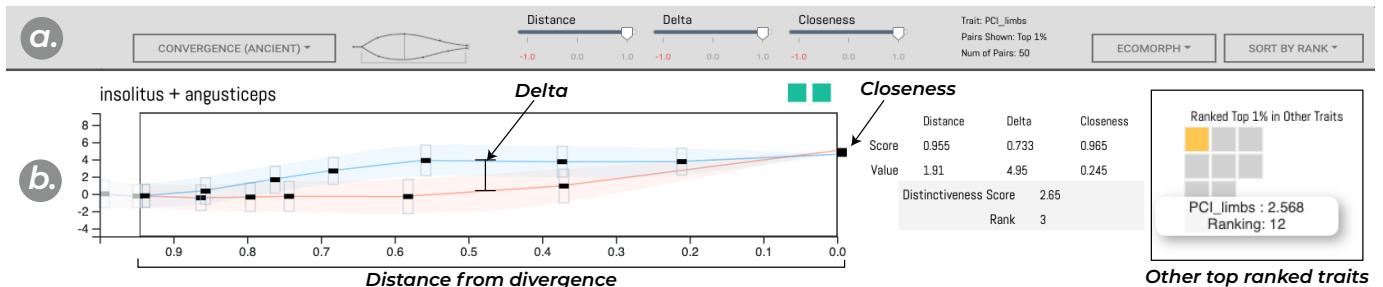


Fig. 4. Pattern view components. (a) The user interface allows selection of a preset pattern, refined by adjusting the parameters for Distance, Delta, and Closeness. This interface also sorts rank pairs by top score or top rank frequency. (b) The first-ranked pair of paths (the species *Anolis insolitus* and *A. angusticeps*) for a convergence pattern for the trait “snout vent length”. The line/area chart shows the most likely values and the associated uncertainty of the trait of consecutive species, with the “delta” between the species in the trait being evident in the middle. Individual species in the two extant species’ ancestry are shown as rectangles. The heat map on the right show where other traits rank based on the selected pattern.

4.6 Conclusions

We developed a web-based visualization tool we call Trevo in collaboration with evolutionary biologists to analyze phylogenetic trees with multivariate and uncertain attributes. In this paper-within-a-paper we contribute two novel visualization techniques implemented in the trait view and the pattern view. The two views prioritize visualizing the attributes in multivariate, phylogenetic trees over detailed topological information. This prioritization is by design. As the tree topology itself is highly uncertain, visualizing uncertain attributes on uncertain nodes is not helpful. Instead, our approach aggregates relevant subtrees by time and visualizes possible attribute distributions for temporal bins. The pattern view similarly prioritizes attributes with only rough topological measures, such as the time two species diverged. It is the first approach that allows researchers to query for complex evolutionary patterns based on a trait and topology, and explore these patterns across multiple traits.

Trevo is being integrated into the computational workflow of the Harmon Lab. Additionally, the development of Trevo is part of a larger software project for creating MultiNet, a web-based tool for visualizing and analyzing multivariate networks [45]. The visualization insights generated from this design study are informing aspects of MultiNet’s design. We discuss the methodological insights generated from this design study in the next section.

5 METHODOLOGICAL RECOMMENDATIONS

Our experiments with design study criteria for rigor — specifically ABUNDANT, REFLEXIVE, and TRANSPARENT — offered us a wealth of opportunities to try new things, and to learn along the way. Through a critically reflective process, we distilled our learning into three methodological recommendations for conducting interpretivist design studies.

5.1 Explicit, systematic reflection is productive

Reflection is a critical aspect of design study [66], yet little is known on how and when to do this in practice [43]. In our work we reflected regularly, and reflexively, documenting our reflections as we progressed through the study. What we found is that systematic reflection shifted the course of our research in productive and demonstrable ways.

For example, when Rogers arrived in our collaborators’ lab at the start of the field study, she initially felt uncomfortable audio-recording her interactions with them. Because she was not familiar with the group and the group was not familiar with her, she felt like an intruder in the lab. In a reflective note from one of her first interviews, she noted:

I have not been recording these interviews as I am in the first week and I do not want to be intrusive. [T3].

She was, however, aware that audio recording would be beneficial to her note taking:

I believe the recording will help me capture more than I can get from my note taking, and maybe more importantly, be more engaged in the interview process. I was initially hesitant to ask people to record them during my initial time

here because I was new and unfamiliar and wanted our first interactions to be more candid. [T53].

The following week she decided to audio-record the participatory workshop she ran with the lab, and reflected on the experience:

I am glad I recorded the workshop — as I have re-listened to it and transcribed parts I felt were significant to the goal of the design study. Returning to the audio at a later time allowed me to notice things that people said when I was engaged in a conversation with someone else or did not have the base knowledge on a particular subject to want to write the moment down initially. [T36]

Rogers’ concerns about her intrusive presence in the lab made her initially hesitant to audio-record interviews, to the detriment of her data gathering. After writing several reflexive memos detailing her feelings, and reflecting on the success of audio-recording the workshop, she changed her interview method and audio-recorded all interviews with collaborators. Off-loading the work of note-taking to the recording allowed her to engage in a more conversational, constructive way when conducting interviews:

I found [audio-recording] extremely helpful as I was able to engage in conversation more easily than when I was attempting to take speed notes....The recording seems to blend into the scene and you forget its running after a couple of minutes. I will be using a recorder from now on. [T53].

By reflecting on her actions, Rogers was able to adjust and improve her research practices. Systematic, reflexive notes such as these are encouraged by qualitative researchers as they offer “a partial means for providing checks on the researcher’s own biases” [35] and a mechanism to “detect and correct deviations from the design goal early” [59].

The start of audio-recording within the design study led to our second example of productive reflection. After conducting an interview, Rogers would listen to the audio-recording from the interview and reflectively transcribe it within a day or two. Transcription did not involve transcribing the audio-recording word for word, but was instead a reflective memo synthesizing the main points taken from the audio along with concrete quotes as evidence for these findings. When something stood out in the recording, Rogers would memo what time in the audio this happened, allowing her to easily revisit how something was said at a later time [T36].

We find that reflectively transcribing an audio-recording — versus relying on an (automatically or externally generated) word-for-word transcription — offers two advantages for analysis. First, listening to the audio while taking notes slows us down, allowing for a deeper, more thoughtful analysis. Writing down reflections requires us to stop, rewind, and listen to things multiple times, resulting in better notes and interpretations. Second, we find that listening to a recording allows us to re-experience the interview, but in a more detached and reflective way. This allows a new perspective on the discussion, separate from the one we experienced in the moment [T126].

Our third example of productive reflection occurred as we constructed an audit trail of our collected artifacts in order to produce a transparent trace of the design study process. Upon revisiting her old sketches, Rogers noticed that her design concepts for certain components were very narrow, particularly for an early version of the trait view [T705, T709, T731]. She reflected on the narrow design concepts during a meeting:

I get fixated on one design and I can see that in the sketches in my sketchbook. [T111].

This reflection prompted Rogers to attempt a redesign of the trait view's discrete plots, which had last gone through design iterations three months prior. Having recently reviewed her notes she took during the field study as she added them to the audit trail, she found new meaning, and new ideas for her redesign:

I still find details that I missed at the time of a meeting or at an initial reflection. [T111].

The redesigned trait view, shown in Figure 2, shows relationships between trait values and their probability distributions that were not shown in earlier designs.

We did not anticipate that the act of curating and organizing artifacts would facilitate productive reflection and play a role in design development. This redesign would likely not have occurred without the reflective processes of revisiting past notes, a concept emphasized in work on systematic reflection for design in engineering. By adopting regular reflection during design, “the chance of overlooking important aspects is decreased” [59]. Tavory and Timmermans advocate for revisiting experiential notes to reconsider them with newfound knowledge or perspective: “We are constantly re-experiencing parts of our world as we go about the business of living. When we move through our surroundings, we not only encounter new problem situations but find new problems in old situations” [74].

RECOMMENDATION Our work shows that adopting regular, systematic, reflective practices within a design study can improve the research methods, domain understanding, and visualization designs. We recommend four opportunities for reflection-on-action. **First**, take reflective notes before and after interviews with domain collaborators. This activity takes only a few minutes but significantly improves the focus of an interview as well as captures initial interpretations and ideas for next steps. **Second**, include reflexive considerations in your field notes. Reflecting on changing perspectives, biases, methodological rationale, and feelings can be a valuable source of insight. **Third**, audio-record interviews and analyze them via reflective transcription. The reflective transcription should occur soon after the interview to support experiential recall on the part of the researcher. **Fourth**, revisit early notes and sketches. During these revisits look for opportunities to reinterpret experiences through a new lens of deeper understanding.

5.2 Traceability supports transparency and reflection

Providing a transparent, scrutinizable trace of a design study is essential for allowing judgments about the quality of the research [44]. As we developed an auditable trace through our collection of research artifacts, we found, however, that revisiting evidence *also* supported productive reflection that shifted and changed the course of the study. Supporting different ways to trace the design study process was important for encouraging both transparency and reflection in our study.

From the start of the field study, we meticulously collected a rich set of research artifacts in order to abundantly document our research process. We stored the artifacts in an online repository, and created a record for each in a spreadsheet that included a descriptive title, the date it was created, a unique id, and the research artifact *type* such as meeting note, sketch, email, etc. Building on this collection of evidence, we experimented with transparent reporting by creating an audit trail of the artifacts. Our initial, web-based design of the audit trail was inspired by those created for other experiments on reporting design studies [29, 40]. Like previous examples, our website traced the design study temporally by visually organizing artifacts on an overview

timeline, and providing access to the recorded artifacts themselves through a details-on-demand side panel. Each artifact is represented on the timeline as a square, color-coded by its type [T161].

While building the audit trail, we reflectively engaged with the research artifacts, leading to demonstrable changes within our study, as we previously discussed in Section 5.1. This engagement shifted the audit trail toward use as an internal, research tool. We found that we wanted to trace research *concepts* across the study, including our growing understanding of domain principles such as convergence and uncertainty, as well as our criteria experiments through reflexivity and sketching. To support concept tracing we extended our metadata for each research artifact to include tags that pull information embedded within the artifacts. These concept tags allow for a trace of how our awareness and understanding of various concepts evolved throughout the study. We extended the website to include the concept tags for each artifact in the detail view; clicking on a specific tag highlights other artifacts with the same tag in the timeline overview [T161].

The final iteration of our tool supports an unanticipated range of research tasks: recording diverse research artifacts, reflecting on conceptual developments, and reporting on the design study process. It is a trace of our research process from two perspectives: a temporal perspective for transparent and auditable reporting and a conceptual perspective for reflective research practices. We consider this tool to be a **ttrrace**, as both a speculative nod to material traces [55] and to the record, reflect, and report tasks it supports.

The ttrrace has theoretical connections to both audit trails [1, 35] and annotated portfolios [8, 21]. As referential material [35], our research artifacts are evidence of the design study process [25, 47, 77, 83], capturing fleeting aspects of the study that led to insight. Organizing these artifacts temporally provides a trace of the study itself [55], providing an auditable mechanism for reviewing the quality of the research [48, 73, 77]. Our research artifacts are also manifestations of design knowledge [21], with the knowledge engrained within the artifact [15]. Each artifact's concept tag, created from the artifact itself, is an annotation, allowing for a trace that connects seemingly disparate artifacts through more general concepts. These theoretical connections point to an opportunity for further theorizing about, and experimenting with, design study ttrraces.

RECOMMENDATION Our experiments with abundant evidence and transparent reporting led us to the concept of a ttrrace, which supports recording, reflecting, and reporting in design study. We recommend that design study researchers plan for a ttrrace early in a study and consider three important issues. First, the process of collecting artifacts greatly benefits from establishing a system for organization early on. We used an online spreadsheet and adopted a regular practice of adding records of digitized artifacts as we generated them. Second, develop mechanisms to automatically extract concept tags from the artifacts themselves. We extracted concepts from the artifacts manually for this project, but in future work we plan to develop an improved, semi-automated approach. Third, the immersive, ethnographic nature of design study requires considerations of how to handle privacy, as well as anonymization for review. We encourage developing a system for anonymizing artifacts early in the study process. Additionally, we find that the best method for navigating transparent recording of a study is to be transparent: tell your collaborators when you are recording, establish what will be on- versus off-record, provide them access to your notes, and be aware of recording delicate social dynamics.

5.3 Methods are necessary, but evidence is the proof

Employing appropriate and justified research methods within a design study is necessary for achieving rigor, but a checklist of methods is not sufficient for arguing that a study is rigorous. The design study rigor criteria are meant to provide guidance on *what* to achieve, not *how* to do so [44]. Evidence of the criteria within a study is the proof. The type, extent, and depth of evidence that is sufficient for arguing that a design study meets the criteria for rigor, however, is an open question, and likely one without a standardizable answer. As part of our experiments we reflected over our research artifacts and experiences,

looking for evidence of the criteria. We found that shifts in the way we communicated and interacted with our collaborators suggest that our study was INFORMED and ABUNDANT.

During the early stages of our field study, work discussions with our collaborators centered around semistructured interviews. We organized interviews to have 2-3 in a single day and scheduled interview days every few days. Rogers saved up questions she had until these interviews [T7, T53]. The infrequent discussions were relatively long in duration, lasting from 1-2 hours at a time, were dense with domain information, and had a formal tone. Post-interviews, Rogers would revisit and look up domain concepts and vocabulary that emerged from these interviews as she was building her understanding of the domain:

The paper [linked] below was a really good resource for getting an understanding of the group comparisons that indicate adaptive events such as convergence.... People I have interviewed touched on these concepts, but because the concepts are complex and varied, it is really hard to get a good synthesis of the main points. I feel as if I am hearing recurring words that come up in conversation, but I have been missing the connection between them. [T87]

As the field study progressed and Rogers felt increasingly comfortable asking questions outside of scheduled interviews, work communications shifted to shorter, informal discussions and texts. The language of communications also shifted as Rogers increasingly used domain vocabulary and concepts fluently. For example, Rogers saw some unexpected biological relationships in the data while developing one of the visualization views, and messaged a collaborator to confirm her observations:

Rogers: *WOOP. Saving the summary view.*

AP: *Booyah*

Rogers: *One thing I noticed the other night is that a lot of the convergent pairs are not both the same ecomorph — but because we are looking at a single trait, would it make sense that two ecomorphs would have similar characteristics for a single trait? Ex: trunk crown and twig having similar PCIII Padwitch vs tail??!*

AP: *Hmmmm.... You're finding that even when using the PC traits? Because those PCs are essentially composites of multiple traits [T133]*

The texts continued as Rogers also excitedly communicated her findings of a problem with the pattern ranking system:

Rogers: *THINK I FOUND A BUG IN THE DELTA.*

AP: *Oh *** What's it doing? [T133]*

At the time of this text exchange, Rogers had spent significant time engaged with the domain, and she understood enough about domain concepts to identify mismatches in what she saw in the data. Furthermore, identifying these mismatches *excited* her.

This exchange aligns with indicators for immersion: using domain-specific language the researcher engages in “informal peer-to-peer communication with domain experts about domain science and visualizations”, eventually becoming “concerned with, affected by, and personally involved in the other domain” [23]. Design by immersion is an approach that, through long-term, committed engagement, provides visualization researchers an abundant exposure to a domain space, allowing them to develop a deeply informed understanding.

Every design study, like other qualitative *inquiries*, is unique in complex ways and thus requires the construction of careful, thoughtful arguments for its quality: “Excellent research is not achieved solely by the use of appropriate strategies or techniques. The skillful use of strategies only sets the stage for the conduct of inquiry” [49]. Changes in the way we communicated with our collaborators — not the time we spent in the field or the number of interviews we conducted — suggests that our design study met aspects of both the INFORMED and ABUNDANT criteria. **We argue for careful argumentation, backed up by rich evidence and grounded in existing literature and theories, as a**

general model for supporting claims of rigor in design studies. Being reflexive and noticing not only how we affected the research, but also how it affected us, offered us opportunities to more deeply reflect on the impacts of our criteria experimentations. We speculate that many such opportunities may be found in any design study.

RECOMMENDATION Knowing when a design study has reached a critical threshold for establishing rigor is difficult, with no single, objective metric. Through critical reflection we positioned our experiences and evidence — shifting patterns of communication — within existing theoretical concepts — design by immersion [23] — allowing us to build links between what occurred in our research and what it could mean. **We recommend that design study researchers plan for the time and space to engage critically and reflectively with their research artifacts and experiences; propose, repropose, and repropose again how what they learned engages with the existing literature; and resist the urge to argue that a study is rigorous because of a checklist of methods employed, instead looking for things that changed, shifted, and surprised.**

6 CONCLUSION

This paper reports on an interpretivist design study and a resulting diverse set of knowledge outcomes consisting of visualization techniques, methodological insights, and new methods for reporting. We found that our experiments with establishing rigor through the ABUNDANT, REFLEXIVE, and TRANSPARENT criteria led to a myriad of learning opportunities. Yet those opportunities are messy, overlapping, and difficult to distill. For example, our efforts to provide transparency relied on abundant data collection, and (reflexively) changed our writing methods as we crafted this paper. We learned much more than we have reported, but the challenge of aligning the evidence, our experience, and existing theory kept us from fully synthesizing the rich learning this interpretivist design study provided.

One such example is the trrrace construct we propose for recording, reflecting, and reporting in design study. The idea of the trrrace emerged as we worked to enhance the transparency of this report. The more we linked into our collection of artifacts, the more we noticed how these links provided useful traces of our research process. We also became aware of challenges for a mechanism like a trrrace that is used in both the research and reporting processes: how do we ensure persistence of the trrrace and the myriad artifacts it links together? How do we consider privacy concerns, as well as anonymization constraints? How do we develop and maintain a trrrace in a way that does not slow down design-oriented research? How do we improve our recording practices to enhance the traceability of a trrrace? How do we report a trrrace in a way that is accessible, understandable, and scrutinizable?

This last question offers opportunities to reflect on current practices for reporting through traditional supplemental materials that can, at best, tell a curated story parallel to a paper, but at worst, can be an impenetrable dump of information. What types of visualizations, interactions, and interfaces can we design to help a reader navigate a trrrace? How might we tell a data-driven story from an abundant collection of evidence? If we embrace the concept of material traces, how might this fundamentally change the way we think about supplemental materials, transparency, and reproducibility? **Developing theory and pragmatic guidance for design study trrraces is one of the more exciting future directions pointed to by this work. We hope this paper is a catalyst for further conversations about trrraces, as well as the broader opportunities and challenges for interpretivist design study.**

ACKNOWLEDGMENTS

We thank the members of the Harmon lab, the Visualization Design Lab, and the MultiNet team for their participation, feedback, and support. We also appreciate the thoughtful comments from the anonymous reviewers that helped to strengthen this paper. We gratefully acknowledge funding by the National Science Foundation (OAC 1835904).

REFERENCES

- [1] S. Akkerman, W. Admiraal, M. Brekelmans, and H. Oost. Auditing quality of research in social sciences. *Quality & Quantity*, 42(2):257–274, 2008. doi: 10.1007/s11135-006-9044-4
- [2] S. Avle and S. Lindtner. Design(ing) “Here” and “There”: Tech Entrepreneurs, Global Markets, and Reflexivity in Design Processes. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI ’16, pp. 2233–2245, 2016. doi: 10.1145/2858036.2858509
- [3] S. Bardzell. Feminist hci: taking stock and outlining an agenda for design. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 1301–1310, 2010. doi: 10.1145/1753326.1753521
- [4] M. Birks, Y. Chapman, and K. Francis. Memoing in qualitative research: Probing data and processes. *Journal of Research in Nursing*, 13(1):68–75, 2008. doi: 10.1177/1744987107081254
- [5] F. Block, M. S. Horn, B. C. Phillips, J. Diamond, E. M. Evans, and C. Shen. The DeepTree Exhibit: Visualizing the Tree of Life to Facilitate Informal Learning. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2789–2798, 2012. doi: 10.1109/TVCG.2012.272
- [6] S. Bødker. When second wave hci meets third wave challenges. In *Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles*, pp. 1–8, 2006. doi: 10.1145/1182475.1182476
- [7] R. R. Bouckaert. DensiTree: Making sense of sets of phylogenetic trees. *Bioinformatics*, 26(10):1372–1373, 2010. doi: 10.1093/bioinformatics/btq110
- [8] J. Bowers. The logic of annotated portfolios: Communicating the value of ‘research through design’. In *Proceedings of the Designing Interactive Systems Conference*, DIS ’12, pp. 68–77, 2012. doi: 10.1145/2317956.2317968
- [9] M. Brehmer, S. Ingram, J. Stray, and T. Munzner. Overview: The design, adoption, and analysis of a visual document mining tool for investigative journalists. *IEEE transactions on visualization and computer graphics*, 20(12):2271–2280, 2014. doi: 10.1109/TVCG.2014.2346431
- [10] S. Brookfield. Critically reflective practice. *Journal of Continuing Education in the Health Professions*, 18(4):197–205, 1998. doi: 10.1002/chp.1340180402
- [11] R. Buchanan. Wicked Problems in Design Thinking. *Design Issues*, 8(2):5–21, 1992. doi: 10.2307/1511637
- [12] B. Buxton. *Sketching User Experiences: Getting the Design Right and the Right Design*. Morgan kaufmann, 2010.
- [13] M. Carcary. The research audit trial—enhancing trustworthiness in qualitative inquiry. *Electronic Journal of Business Research Methods*, 7(1), 2009.
- [14] N. Cross. Design research: A disciplined conversation. *Design issues*, 15(2):5–10, 1999. doi: 10.2307/1511837
- [15] N. Cross. Expertise in design: an overview. *Design Studies*, 25(5):427–441, 2004. doi: 10.1016/j.destud.2004.06.002
- [16] A. Daley. Reflections on Reflexivity and Critical Reflection as Critical Research Practices. *Affilia*, 25(1):68–82, 2010. doi: 10.1177/0886109909354981
- [17] R. De Kleijn and A. Van Leeuwen. Reflections and review on the audit procedure: Guidelines for more transparency. *International Journal of Qualitative Methods*, 17(1), 2018. doi: 10.1177/1609406918763214
- [18] S. P. Dow, A. Glassco, J. Kass, M. Schwarz, D. L. Schwartz, and S. R. Klemmer. Parallel prototyping leads to better design results, more divergence, and increased self-efficacy. *ACM Transactions on Computer-Human Interaction*, 17(4):18:1–18:24, 2011. doi: 10.1145/1879831.1879836
- [19] L. Finlay. Negotiating the swamp: the opportunity and challenge of reflexivity in research practice. *Qualitative research*, 2(2):209–230, 2002. doi: 10.1177/146879410200200205
- [20] C. Frayling. Research in art and design. *Royal College of Art Research Papers*, 1(1), 1994.
- [21] B. Gaver and J. Bowers. Annotated portfolios. *Interactions*, 2012.
- [22] J. Guerra-Gomez, M. L. Pack, C. Plaisant, and B. Shneiderman. Visualizing Change Over Time Using Dynamic Hierarchies: TreeVersity2 and the StemView. *IEEE Transactions on Visualization and Computer Graphics (InfoVis ’13)*, 19(12):2566–2575, 2013.
- [23] K. W. Hall, A. J. Bradley, U. Hinrichs, S. Huron, J. Wood, C. Collins, and S. Carpendale. Design by Immersion: A Transdisciplinary Approach to Problem-Driven Visualizations. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2019. doi: 10.1109/TVCG.2019.2934790
- [24] D. Haraway. Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist studies*, 14(3):575–599, 1988. doi: 10.2307/3178066
- [25] B. Hartmann, M. R. Morris, H. Benko, and A. D. Wilson. Pictionnaire: Supporting collaborative design work by integrating physical and digital artifacts. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, CSCW ’10, pp. 421–424, 2010. doi: 10.1145/1718918.1718989
- [26] U. Hinrichs, S. Forlini, and B. Moynihan. Speculative practices: Utilizing infovis to explore untapped literary collections. *IEEE transactions on visualization and computer graphics*, 22(1):429–438, 2015. doi: 10.1109/TVCG.2015.2467452
- [27] D. H. Huson and C. Scornavacca. Dendroscope 3: An Interactive Tool for Rooted Phylogenetic Trees and Networks. *Systematic Biology*, 61(6):1061–1067, 2012. doi: 10.1093/sysbio/sys062
- [28] T. Ingram and D. L. Mahler. SURFACE: Detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike Information Criterion. *Methods in Ecology and Evolution*, 4(5):416–425, 2013. doi: 10.1111/2041-210X.12034
- [29] E. Kerzner, S. Goodwin, J. Dykes, S. Jones, and M. Meyer. A Framework for Creative Visualization-Opportunities Workshops. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):748–758, 2019. doi: 10.1109/TVCG.2018.2865241
- [30] L. F. Klein and C. D’Ignazio. *Data feminism*. MIT Press, 2020.
- [31] F. A. J. Korthagen, J. Kessels, B. Koster, B. Lagerwerf, and T. Wubbels. *Linking practice and theory: The pedagogy of realistic teacher education*. Lawrence Erlbaum Associates Publishers, 2001.
- [32] P.-M. Law, R. C. Basole, and Y. Wu. Duet: Helping data analysis novices conduct pairwise comparisons by minimal specification. *IEEE transactions on visualization and computer graphics*, 25(1):427–437, 2018. doi: 10.1109/TVCG.2018.2864526
- [33] B. Lee, G. G. Robertson, M. Czerwinski, and C. S. Parr. CandidTree: Visualizing Structural Uncertainty in Similar Hierarchies. In C. Baranauskas, P. Palanque, J. Abascal, and S. D. J. Barbosa, eds., *Human-Computer Interaction – INTERACT 2007*, Lecture Notes in Computer Science, pp. 250–263, 2007. doi: 10.1007/978-3-540-74800-7_20
- [34] I. Letunic and P. Bork. Interactive Tree Of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Research*, 47(W1):W256–W259, 2019. doi: 10.1093/nar/gkz239
- [35] Y. S. Lincoln and E. G. Guba. Establishing Dependability and Confirmability in Naturalistic Inquiry Through an Audit. In *Annual Meeting of the American Educational Research Association*, 1982.
- [36] Z. Liu, S. H. Zhan, and T. Munzner. Aggregated Dendograms for Visual Comparison Between Many Phylogenetic Trees. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2019. doi: 10.1109/TVCG.2019.2898186
- [37] D. Macbeth. On “Reflexivity” in Qualitative Research: Two Readings, and a Third. *Qualitative Inquiry*, 2016. doi: 10.1177/107780040100700103
- [38] S. Mantzoukas. The inclusion of bias in reflective and reflexive research: A necessary prerequisite for securing validity. *Journal of Research in Nursing*, 10(3):279–295, 2005. doi: 10.1177/174498710501000305
- [39] N. McCurdy, J. Dykes, and M. Meyer. Action Design Research and Visualization Design. In *Proceedings of the Beyond Time and Errors on Novel Evaluation Methods for Visualization - BELIV ’16*, pp. 10–18, 2016. doi: 10.1145/2993901.2993916
- [40] N. McCurdy, J. Gerdes, and M. Meyer. A Framework for Externalizing Implicit Error Using Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):925–935, 2019. doi: 10.1109/TVCG.2018.2864913
- [41] N. McCurdy, J. Lein, K. Coles, and M. Meyer. Poemage: Visualizing the sonic topology of a poem. *IEEE transactions on visualization and computer graphics*, 22(1):439–448, 2015. doi: 10.1109/TVCG.2015.2467811
- [42] S. McKenna, D. Mazur, J. Agutter, and M. Meyer. Design Activity Framework for Visualization Design. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2191–2200, 2014. doi: 10.1109/TVCG.2014.2346331
- [43] M. Meyer and J. Dykes. Reflection on reflection in applied visualization research. *IEEE Computer Graphics and Applications*, 38(6):9–16, 2018. doi: 10.1109/MCG.2018.2874523
- [44] M. Meyer and J. Dykes. Criteria for Rigor in Visualization Design Study. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2019. doi: 10.1109/TVCG.2019.2934539
- [45] M. Meyer, A. Lex, J. Baumes, C. Lisle, and L. Harmon. Collaborative research: Framework: Software: Hdr: Reproducible visual analysis of

- multivariate networks with MultiNet, 2019. <https://vd1.sci.utah.edu/projects/2019-nsf-multinet/>.
- [46] M. Meyer, M. Sedlmair, P. S. Quinan, and T. Munzner. The nested blocks and guidelines model. *Information Visualization*, 14(3):234–249, 2015. doi: 10.1177/1473871613510429
- [47] B. Middleton, J. Platt, J. Richardson, and B. Blumenfeld. Recommendations for Building and Maintaining Trust in Clinical Decision Support Knowledge Artifacts. *Research Triangle Park, NC: Patient-Centered Clinical Decision Support Learning Network*, 2018.
- [48] A. Moravcsik. Transparency: The Revolution in Qualitative Research. *PS: Political Science & Politics*, 47(1):48–53, 2014. doi: 10.1017/S1049096513001789
- [49] J. Morse. The changing face of qualitative inquiry. *International Journal of Qualitative Methods*, 19:1–7, 2020. doi: 10.1177/1609406920909938
- [50] T. Munzner. A Nested Model for Visualization Design and Validation. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):921–928, 2009. doi: 10.1109/TVCG.2009.111
- [51] T. Munzner, F. Guimbretière, S. Tasiran, L. Zhang, and Y. Zhou. TreeJuxtaposer: Scalable Tree Comparison Using Focus+Context with Guaranteed Visibility. In *Proceedings of the ACM Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '03)*, pp. 453–462, 2003. doi: 10.1145/1201775.882291
- [52] C. Nobre, N. Gehlenborg, H. Coon, and A. Lex. Lineage: Visualizing Multivariate Clinical Data in Genealogy Graphs. *Transaction on Visualization and Computer Graphics*, 25(3):1543–1558, 2019. doi: 10.1109/TVCG.2018.2811488
- [53] C. Nobre, M. Meyer, M. Streit, and A. Lex. The State of the Art in Visualizing Multivariate Networks. *Computer Graphics Forum (EuroVis)*, 38(3):807–832, 2019. doi: 10.1111/cgf.13728
- [54] C. Nobre, M. Streit, and A. Lex. Juniper: A Tree+Table Approach to Multivariate Graph Visualization. *Transaction on Visualization and Computer Graphics (InfoVis '18)*, 25(1):544–554, 2019. doi: 10.1109/TVCG.2018.2865149
- [55] D. Offenhuber. Data by Proxy – Material Traces as Autographic Visualizations. *arXiv:1907.05454 [cs, math]*, 2019.
- [56] C. Partl, S. Gratzl, M. Streit, A. M. Wassermann, H. Pfister, D. Schmalstieg, and A. Lex. Pathfinder: Visual Analysis of Paths in Graphs. *Computer Graphics Forum (EuroVis '16)*, 35(3):71–80, 2016. doi: 10.1111/cgf.12883
- [57] L. J. Revell. Phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3(2):217–223, 2012. doi: 10.1111/j.2041-210X.2011.00169.x
- [58] L. J. Revell. Two new graphical methods for mapping trait evolution on phylogenies. *Methods in Ecology and Evolution*, 4(8):754–759, 2013. doi: 10.1111/2041-210X.12066
- [59] I. M. M. J. Reymen and D. K. Hammer. Structured Reflection for Improving Design Processes. In *DS 30: Proceedings of DESIGN 2002, the 7th International Design Conference, Dubrovnik*, pp. 887–892, 2002.
- [60] E. L. Rezende and J. A. F. Diniz-Filho. Phylogenetic Analyses: Comparing Species to Infer Adaptations and Physiological Mechanisms. In *Comprehensive Physiology*, pp. 639–674. Wiley Online Library, 2012. doi: 10.1002/cphy.c100079
- [61] J. A. Rode. Reflexivity in digital anthropology. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11*, pp. 123–132, 2011. doi: 10.1145/1978942.1978961
- [62] J. Rosindell and L. J. Harmon. OneZoom: A Fractal Explorer for the Tree of Life. *PLOS Biology*, 10(10):e1001406, 2012. doi: 10.1371/journal.pbio.1001406
- [63] A. Satyanarayan, B. Lee, D. Ren, J. Heer, J. Stasko, J. Thompson, M. Brehmer, and Z. Liu. Critical reflections on visualization authoring systems. *IEEE transactions on visualization and computer graphics*, 26(1):461–471, 2019. doi: 10.1109/TVCG.2019.2934281
- [64] D. A. Schon. *The Reflective Practitioner: How Professionals Think In Action*. Basic books, 1984.
- [65] M. Sedlmair. Design study contributions come in different guises: Seven guiding scenarios. In *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*, pp. 152–161, 2016. doi: 10.1145/2993901.2993913
- [66] M. Sedlmair, M. Meyer, and T. Munzner. Design Study Methodology: Reflections from the Trenches and the Stacks. *IEEE Transactions on Visualization and Computer Graphics (InfoVis)*, 18(12):2431–2440, 2012. doi: 10.1109/TVCG.2012.213
- [67] M. K. Sein, O. Henfridsson, S. Purao, M. Rossi, and R. Lindgren. Action design research. *MIS quarterly*, pp. 37–56, 2011. doi: 10.2307/23043488
- [68] A. K. Shenton. Strategies for ensuring trustworthiness in qualitative research projects. *Education for information*, 22(2):63–75, 2004. doi: 10.3233/EFI-2004-22201
- [69] S. Simon, S. Mittelstädt, D. A. Keim, and M. Sedlmair. Bridging the gap of domain and visualization experts with a liaison. In *Accepted at the Eurographics Conference on Visualization (EuroVis 2015, Short Paper)*, vol. 2015. The Eurographics Association, 2015. doi: 10.2312/eurovisshort.20151137
- [70] B. Smith and K. R. McGannon. Developing rigor in qualitative research: Problems and opportunities within sport and exercise psychology. *International review of sport and exercise psychology*, 11(1):101–121, 2018. doi: 10.1080/1750984X.2017.1317357
- [71] P. J. Stappers. Doing design as a part of doing research. In *Design research now*, pp. 81–91. Springer, 2007. doi: 10.1007/978-3-7643-8472-2_6
- [72] U. H. Syeda, P. Murali, L. Roe, B. Berkey, and M. A. Borkin. Design Study "Lite" Methodology: Expediting Design Studies and Enabling the Synergy of Visualization Pedagogy and Social Good. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20*, pp. 1–13, 2020. doi: 10.1145/3313831.3376829
- [73] P. Talkad Sukumar, I. Avellino, C. Remy, M. A. DeVito, T. R. Dillahunt, J. McGrenere, and M. L. Wilson. Transparency in Qualitative Research: Increasing Fairness in the CHI Review Process. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, CHI EA '20*, pp. 1–6, 2020. doi: 10.1145/3334480.3381066
- [74] I. Tavory and S. Timmermans. *Abductive Analysis: Theorizing Qualitative Research*. University of Chicago Press, 2014.
- [75] S. Thompson and N. Thompson. *The critically reflective practitioner*. Macmillan International Higher Education, 2018.
- [76] S. J. Tracy. Qualitative quality: Eight big-tent criteria for excellent qualitative research. *Qualitative inquiry*, 16(10):837–851, 2010. doi: 10.1177/1077800410383121
- [77] C. Wacharamanottham, L. Eisenring, S. Haroz, and F. Echtler. Transparency of CHI Research Artifacts: Results of a Self-Reported Survey. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20*, pp. 1–14, 2020. doi: 10.1145/3313831.3376448
- [78] Y. L. Wong, K. Madhavan, and N. Elmqvist. Towards characterizing domain experts as a user group. In *2018 IEEE Evaluation and Beyond-Methodological Approaches for Visualization (BELIV)*, pp. 1–10. IEEE, 2018. doi: 10.1109/BELIV.2018.8634026
- [79] J. Wood, A. Kachkaev, and J. Dykes. Design exposition with literate visualization. *IEEE transactions on visualization and computer graphics*, 25(1):759–768, 2018. doi: 10.1109/TVCG.2018.2864836
- [80] D. Yanow and H. Tsoukas. What is Reflection-In-Action? A Phenomenological Account. *Journal of Management Studies*, 46(8):1339–1364, 2009. doi: 10.1111/j.1467-6486.2009.00859.x
- [81] G. Yu. Using ggtree to Visualize Data on Tree-Like Structures. *Current Protocols in Bioinformatics*, 69(1):e96, 2020. doi: 10.1002/cpb1.96
- [82] Y. Zhang, K. Chanana, and C. Dunne. Idmviz: Temporal event sequence visualization for type 1 diabetes treatment decision support. *IEEE transactions on visualization and computer graphics*, 25(1):512–522, 2018. doi: 10.1109/TVCG.2018.2865076
- [83] J. Zimmerman and J. Forlizzi. The Role of Design Artifacts in Design Theory Construction. *Artifact*, 2(1):41–45, 2008. doi: info:doi/10.1080/17493460802276893
- [84] J. Zimmerman, J. Forlizzi, and S. Evenson. Research through design as a method for interaction design research in hci. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 493–502, 2007. doi: 10.1145/1240624.1240704