

Video Collaboratories for Research and Education: An Analysis of Collaboration Design Patterns

Roy Pea, *Member, IEEE*, and Robb Lindgren, *Student Member, IEEE*

Abstract—Web-based video collaboration environments have transformative potentials for video-enhanced education and for video-based research studies. We first describe DIVER, a platform designed to solve a set of core challenges we have identified in supporting video collaboratories. We then characterize five Collaboration Design Patterns (CDPs) that emerged from numerous collaborative groups who appropriated DIVER for their video-based practices. Collaboration Design Patterns (CDPs) are ways of characterizing interaction patterns in the uses of collaboration technology. Finally, we propose a three-dimensional design matrix for incorporating these observed patterns. This representation can serve heuristically in making design suggestions for expanding video collaboratory functionalities and for supporting a broader constellation of user groups than those spanned by our observed CDPs.

Index Terms—Video, collaboration, CSCL, collaboratory, design patterns, Web 2.0, metadata, video analysis.

1 INTRODUCTION

WE argue in this paper that the proliferation of digital video recording, computing, and Internet communications in the contexts of social sciences research and learning technologies has opened up dramatic new possibilities for creating “video collaboratories.” Before describing our vision for video collaboratories and our experiences in designing and implementing the DIVER platform for enabling video collaboration, we briefly sketch the historical developments in video use leading to the opportunities at hand.

Throughout the 20th century, film (and, later, video) technology was an influential medium for capturing rich multimedia records of the physical and social worlds for educational purposes [1] and for research uses in the social sciences [2]. While K-20 education is still largely dominated by textual representations of information and uses of static graphics and diagrams, during the 21st century the education and research communities have more regularly exploited video technologies, and in increasingly innovative ways. For example, with the development of more learner-centered pedagogy, uses of video are expanding from teachers simply showing videos to students to approaches where learners interact with, create, or comment on video resources as part of their knowledge-building activities [3], [4], [5], [6]. In research, with the proliferation of inexpensive digital consumer video cameras and software for video editing and analysis, individual researchers and research teams are capturing more data for studying the contextual

details of learning and teaching processes, and the learning sciences community has begun experimenting with collaborative research infrastructures surrounding video data sets [7], [8], [9], [10].

We view the Web 2.0 participatory media culture illustrated by media-sharing community sites [11] as exemplifying how new forms of collaboration and communication have important transformative potentials for more deeply engaging the learner in authentic forms of learning and assessment that get closer to the experiences of worldly participation rather than more traditional decontextualized classroom practices. Video representations provide a medium of great importance in these transformations in capturing the everyday interactions between people in their physical environments during their engagements in cultural practices and using the technologies that normally accompany them in their movement across different contexts. For these reasons, video is important both as a medium for studies of learning and human interaction and for educational interventions. In research, the benefits of video have been well evidenced. Learning scientists have paid increasing attention over the past two decades to examining human activities in naturalistic sociocultural contexts, with an expansion of focus from viewing learning principally as an internal cognitive process toward a view of learning that is also constituted as a complex social phenomenon involving multiple agents, symbolic representations, and environmental features and tools to make sense of the world and one another [12], [13].

This expansion in the central focus for studies of learning, thinking, and human practices was deeply influenced by contributions involving close analyses of video and audio recordings from conversation analysis, sociolinguistic studies of classroom discourse, anthropological and ethnographic inquiries of learning in formal and informal settings, and studies of socially significant non-verbal behaviors such as “body language” or kinesics,

• The authors are with the H-STAR Institute and the School of Education, Stanford University, Wallenberg Hall, Building 160, 450 Serra Mall, Stanford, CA 94305-2055. E-mail: {roypea, robblind}@stanford.edu.

Manuscript received 9 Dec. 2008; revised 5 Jan. 2009; accepted 7 Jan. 2009; published online 14 Jan. 2009.

For information on obtaining reprints of this article, please send e-mail to: lt@computer.org, and reference IEEECS Log Number TLT-2008-12-0106. Digital Object Identifier no. 10.1109/TLT.2009.5.

gesture, and gaze patterns. This orientation to understanding learning *in situ* has led researchers to search for tools allowing for the capture of the complexity of real life learning situations, where multiple simultaneous “channels” of interaction are potentially relevant to achieving a deeper understanding of learning behavior [2]. Uses of film and audio-video recordings have been essential in allowing for the repeated and detailed analyses that researchers have used to develop new insights about learning and cultural practices [14], [15]. In research labs throughout departments of psychology, education, sociology, linguistics, communication, (cultural) anthropology, and human-computer interaction, researchers work individually or in small collaborative teams—often across disciplines—for the distinctive insights that can be brought to the interpretation and explanation of human activities using video analysis. Yet, there has been relatively little study of how distributed groups make digital video analysis into a collaborative enterprise, nor have there been tools available that effectively structure and harvest collective insights.

We are inspired by the remarkable possibilities for establishing “video collaboratories” for research and for educational purposes [7], [16]. In research-oriented video collaboratories, scientists will work together to share video data sets, metadata schemes, analysis tools, coding systems, advice and other resources, and build video analyses together, in order to advance the collective understanding of the behaviors represented in digital video data. Virtual repositories with video files and associated metadata will be stored and accessed across many thousands of federated computer servers. A large variety of types of interactions are increasingly captured in video data, with important contexts including K-20 learning—as in ratio and proportion in middle school mathematics or college reasoning about mechanics, parent-child or peer-peer situations in informal learning, surgery and hospital emergency rooms and medical education, aircraft cockpits or other life-critical control centers, focus group meetings or corporate workgroups, deaf sign language communications, and uses of various products in their everyday environments to help guide new design (including cars, computers, cellphones, household appliances, medical devices), and so on. Corresponding opportunities exist for developing education-centered video collaboratories for the purposes of technology-enhanced learning and teaching activities that build knowledge exploiting the fertile properties we have mentioned of audio-video media. It is our belief that enabling scientific and educational communities to develop flexible and sustained interactions around video analysis and interpretation will help accelerate advances across a range of disciplines, as the development of their collective intelligence is facilitated.

We recognize how this vision of widespread digital video collaboratories used throughout communities for research and for education presents numerous challenges. The process of elucidating and addressing these challenges can be aided considerably by exploring emerging efforts to support collaborative video practices. In this paper, we describe the features of a particular digital video collaboratory known as DIVER. Using the large volume of data that

we have collected from DIVER users, we are able to describe the substantial challenges associated with establishing collaboration around digital video using examples from real-world research and educational practices. This data set also permits us to extrapolate future collaboration possibilities and the new challenges they create. We end by presenting dimensions for organizing our vision of digital video collaboratories that we hope will provide entry points for researchers and designers to engage in its further realization.

2 DIVER: DIGITAL INTERACTIVE VIDEO EXPLORATION AND REFLECTION

2.1 The Need for Supporting Video Conversations

We can distinguish three genres of video in collaboration. The first is *videoconferencing*, which establishes synchronous virtual presence (ranging from Skype/iChat video on personal computers to dedicated room-based videoconferencing systems such as HP’s Halo)—where video is the medium, as collaboration occurs via video. The second is *video cocreation* (e.g., Kaltura, Mediasilo)—where video is the objective, and the collaboration is about making video. The third is *video conversations*—where video is the content, and the collaboration is about the video. We feel that video conversations are a vital video genre for learning and education because conversational contributions about videos often carry content as or more important than the videos themselves—the range of interpretations and connections made by different people, which provides new points of view [8], [9] and generates important conceptual diversity [17]. For decades, video has been broadcast-centric—consider TV, K-12 education or corporate training films, and e-learning video. But with the growth of virtual teams, we need a multimediated collaboration infrastructure for sharing meaning and iterative knowledge building across multiple cultures and perspectives. We need a video infrastructure that is more interaction-centric—for people to communicate deeply, precisely, and cumulatively about the video content.

In our vision of video collaboratories, effectively supporting video conversations requires more than the capabilities of videoconferencing and net meetings. One requirement concerns a method for pointing to and annotating parts of videos—analogue to footnoting for text—where the scope of what one is referring to can be made readily apparent. We are beginning to see this capability emerge with interactive digital video applications online. In June 2008, YouTube enabled users to mark a spotlight region and create a pop-up annotation with a start and end time in the video stream. Flash note overlays on top of video streams are also provided in the popular Japanese site Nico Nico Douga, launched in December 2006, where users can post a chat message at a specific moment in the video and other chat messages other users have entered at that time point in the video stream together across the video as it plays. Similar capabilities of “deep tagging” of video were illustrated in the past few years by BubblePly, Click.tv, Eyespot, Gotoit, Jumpcut, Motionbox,

The screenshot shows the DIVER web interface. At the top, it says "Diver™" and "Welcome FM_Researcher 1, Log-out". Below this are navigation links: "Dives | Movies | My Diver | Upload | Signup | Help | About". On the right, there are utility links: "save as | paste | clipboard | remix".

The main content area shows a video player with a "virtual viewfinder" (a white rectangle) overlaid on a video of people sitting at a table. Below the video is a timeline with "MARK" and "RECORD" buttons. The text "Currently playing: ah 1 1" is visible.

Below the video player, there are two recorded segments:

- 1) 19:53:08 <--> 20:18:00** with "share | free-D | copy" links. A thumbnail shows a person pointing at a board. The annotation reads: "In this sequence the father is discussing a geometry problem he faced when constructing a new fence." It was posted by FM_Researcher 1 on Wed Jun 28 09:34:09 PM 2006. There is an "Add comment (0)" link.
- 2) 19:55:13** with "share | free-D | copy" links. A thumbnail shows a person framing a space. The annotation reads: "Here he frames his gesture space for the problem he is about to describe." It was posted by FM_Researcher 2 on Wed Jun 28 09:36:41 PM 2006. There is an "Add comment (0)" link.

Fig. 1. The DIVER user interface. The rectangle inside the video window represents a virtual viewfinder that is controlled by the user's mouse. Users essentially make a movie inside the source movie by recording these mouse movements.

Mojiti (acquired by News Corporation/CBS joint venture Hulu), Veotag, and Viddler.

While the virtual pointing requirement is necessary for supporting online video conversations, it is not sufficient. It is important to distinguish between simple annotation and conversation—the former can be accomplished with tools for coupling text and visual referents, while the latter requires additional mechanisms for managing conversational turn-taking and ensuring multiparty engagement with the target content. While there are numerous software environments currently supporting video annotation, the number of platforms that support video conversations is much smaller.¹ We focus here on a software platform called DIVER that was developed in our Stanford lab with the objective of supporting video conversations. In addition to providing a unique method for pointing and annotating, DIVER also possesses functionality for facilitating and integrating multiuser contributions.

2.2 DIVER as an Example of a Video Conversations Platform

DIVER is a software environment first developed as a desktop software system for exploring research uses of panoramic video records that encompass 360-degree imagery from a dynamic visual environment such as a classroom or a professional meeting [19]. The Web version of the DIVER platform in development and use since 2004 allows a user to control a "virtual camera window" overlaid

1. *VideoTraces* [18] is one system that we are familiar with that strives to support rich asynchronous discourse with video using gestural and voice annotations in a threaded format.

on a standard video record streamed through a Web browser such that the user can "point" to the parts of the video they wish to highlight (Fig. 1). The user can then associate text annotations with the segments of the video being referenced and publish these annotations online so that others can experience the user's perspective and respond with comments of their own. In this way, DIVER enables creating an infinite number of new digital video clips and remix compilations from a single source video recording. As we have modified DIVER to allow distributed access for viewing, annotation, commentary, and remixing, our focus has shifted to supporting collaborative video analysis and the emerging prospects for digital video collaboratories. DIVER and its evolving capabilities have been put to work in support of collaborative video analysis for a diverse range of research and educational activities, which we characterize in the next section.

We refer to the central work product in DIVER as a "dive" (as in "dive into the video"). A dive consists of a set of XML metadata pointers to segments of digital video stored in a database and their associated text annotations. In authoring dives on streaming videos via any Web browser, a user is directing the attention of others who view the dive to see what the author sees; it is a process we call "guided noticing" [16], [19]. To author a dive with DIVER, a user logs in and chooses any video record in the searchable database that they have permission to access (according to the groups to which they belong). A dive can be constructed by a single user or by multiple users, each of whom contributes their own interpretations that may build on the interpretations of the other users.

You've been invited to watch the following collection of stills, movie clips, and commentary....

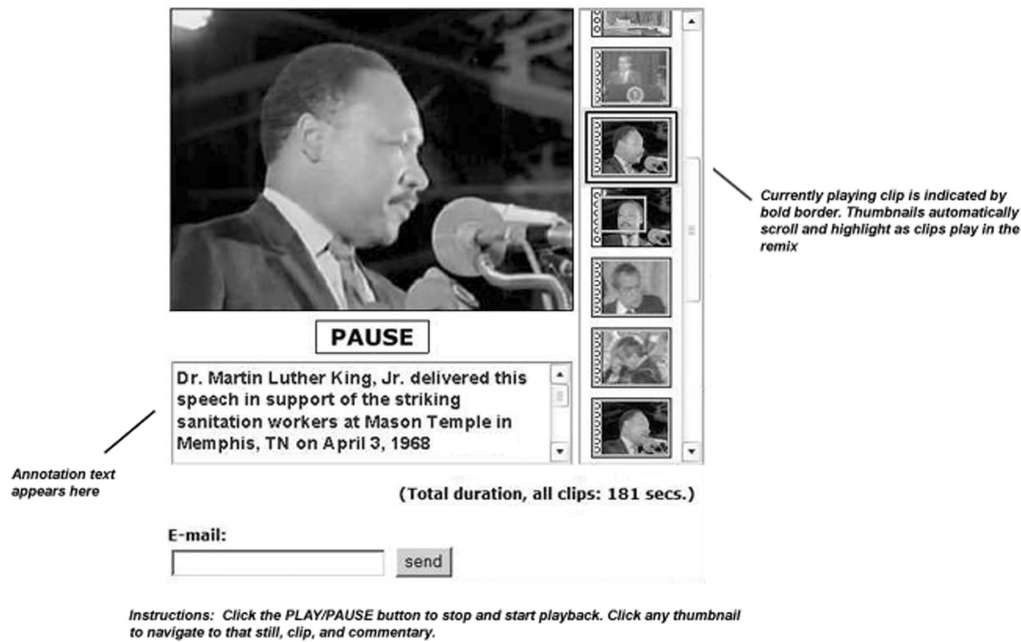


Fig. 2. DIVER remix player for viewing dive content outside of the authoring environment (on a Webpage or blog).

By clicking the “Mark” button in the DIVER interface (see Fig. 1), the user saves a reference to a specific point in space and time in the video. The mark is represented by a thumbnail image of the video record within the DIVER “worksheet” on the right side of the interface. Once the mark is added to the worksheet, the user can annotate that mark by entering text in the associated panel. A panel is also created by clicking on the “Record” button, an action that creates a pointer to a temporal video segment and the video content encompassed within the virtual viewfinder path during that segment. Like a mark, a recorded clip can be annotated by adding text within its associated DIVER worksheet panel. The DIVER user can replay the recorded video path movie or access the point in the video referenced in a mark by clicking on its thumbnail image. DIVER is unique among deep video tagging technologies in enabling users to create and annotate panned and zoomed path movies within streaming Web video.

In addition to asynchronous collaborative video analysis using DIVER, multiple users can simultaneously access a dive, with each user able to add new panels or make comments on a panel that another user has created. Users are notified in real time when another user has made a contribution. Thus users may be either face-to-face in a meeting room or connected to the same Webpage remotely via networking as they build a collaborative video analysis. There is no need for the users to be watching the same portions of the video at the same time. As the video is streamed through their browsers, users may mark and record and comment at their own pace and according to their own interests.

The DIVER user may also create a compilation or “remix” of the contents of the dive as a stand-alone

presentation and share this with collaborators by sending email with a live URL link that will open a DIVER server page (see Fig. 2) and display a DIVER player for viewing the dive including its annotations.

The current architecture of WebDIVER has three parts: 1) a video transcoding server (Windows XP and FFmpeg), 2) application servers that are WAMP XP-based (Windows XP, Apache Web Server, MySQL database for FLV-format videos, thumbnail images, XML files and logs, and PHP), and 3) the streaming media server (Flash Media Server 2, Windows Server 2003). Progress is underway in moving WebDIVER into Amazon Elastic Compute Cloud (Amazon EC2), using Amazon’s S3 storage service, converting to Linux, and replacing Adobe Flash Media Server with open-source Red5 so that we can have a 100-percent open source DIVER.

2.3 Sociotechnical Design Challenges for Video Collaboratories

The Web version of DIVER was designed specifically to address a set of core problems in supporting the activities of participants involved in distributed video collaborations [20]. These problems have surfaced in various multi-institutional workshops convening video researchers [21], [22], [23], [15], and many of the concerns relate to the fundamental problem of coordination of attention, interpretation, and action between multiple persons. Clark [24], in characterizing non-technology-mediated communication, described as “common ground” what people seek to achieve as they coordinate what they are attending to and/or referring to, so that when comments are made, what these comments refer to can be appropriately inferred or elaborated. We expand the common ground aspect of communicative coordination to the need to refer to specific

states of information display in using computer tools, including digital video in collaborative systems.

Of course, a video conversation system such as DIVER does not fully resolve the common ground problem anymore than pointing does in the physical world. While pointing is a useful way of calling attention to something, as Clark and others, such as Quine [25], have pointed out, there can still be referential ambiguity present in a pointing act even for face-to-face discourse (e.g., is she referring to the car as an object or the color of the car?). However, it is possible to design a software platform with functionality that makes it possible for participants to *negotiate the identity of a referent and its meaning* over progressive conversational turns. Marking points of interest in a video using DIVER's virtual viewfinder and creating persistent references to video events are important steps forward in addressing these coordination challenges. A user who offers an interpretation on a movie moment in a DIVER worksheet can be confident that the referent of their comments—at least, in terms of spatial and temporal position—will be unambiguous to others who view the dive. This “guided noticing” feature provides the basis for subsequent interpretive work by group members to align their loci of interpretive attention (e.g., What is this? What is happening here?). In the next section, we examine the effects of introducing this capability to users with varying interests and goals for working with video.

We summarize here the core sociotechnical design challenges we addressed in the design of DIVER [20]:

1. *The problem of reference.* When analyzing video, how does one create a lasting reference to a point in space and time in the dynamic, time-based medium of a video record (a “video footnote”)?
2. *The problem of attentional alignment, or coreference.* How does one know that what one is referring to and annotating in a video record is recognized by one's audience? Such coordination is important because dialog about a video referent can lead to conversational troubles if one's audience has a different referent in mind. This process of deixis requires a shared context.
3. *The problem of effective search, retrieval, and experiencing of collaborative video work.* If we can solve the problems of allowing users to point to video and express an interpretation and to establish attentional alignment, new digital objects proliferate beyond the video files themselves (“dives”) that need to be searchable and readily yield the users' experience of those video moments that matter to them.
4. *The problem of permissions.* How does one make the sharing of video and its associated interpretations more available while maintaining control over sensitive data that should have protection for human subjects or digital rights management?
5. *Integrating the insights of a collective group.* How can we support harvesting and synthesizing the collective intelligence of a group collaboratively analyzing video?
6. *The problem of establishing coherent multiparty video-anchored discourse.* Consider a face-to-face conversational interaction, as in a seminar, where the rules of

discourse that sustain turn-taking and sense-making as people converse are familiar. In an academic setting, video, film, and audio recordings, as well as paper, may be used. Traditionally, these are used asymmetrically, as the facilitator/instructor prepares these records to make a point or to serve as an anchor for discussion and controls their play during the discourse. Computer-facilitated meetings for doing video analysis—where each participant has a computer and is network-connected with other participants and to external networks for information access—bring new challenges in terms of managing a coherent group discourse.

3 COLLABORATION PATTERNS IN DIVER

We now investigate what collaboration design patterns are used by groups of educators and researchers when they have access to DIVER's Web-based video collaboration platform. The DIVER software platform allows users to flexibly exchange analyses and interpretations of a video record with relatively few constraints on their collaboration style or purpose. Once the DIVER software was Web-enabled and made publicly accessible in 2004, we took the approach of supporting any user with a need for engaging in video exploration as part of their professional, educational, or recreational practices. We sought to accommodate a range of video activities, but rarely did we actively recruit users or make customizations to support specific applications. The result has been a large and diverse international user base. There have been approximately 3,000 unique users who have registered with DIVER and approximately 200 private user groups that have been created to support the activities of a particular project, event, or organization. Since DIVER was created in an academic setting, the majority of users are affiliated with schools and universities; however, there have also been users from the private sector, and even within a single setting, we have found DIVER to be used for a number of different purposes.

To conduct this analysis, we examined every dive that was created by each of the user groups and characterized the way that each group had organized their discourse: what interface features were used most frequently, what annotation conventions the group adopted, how conversational turn-taking was managed, etc. These characterizations were agnostic as to who the particular participants were in the collaboration (e.g., high school students or researchers), and we regularly observed behaviors that transcended such user categories. In reviewing these characterizations, we were able to extract a small set of distinct patterns that describe the ways groups collaborated to achieve their particular video learning objectives. We think of these patterns as akin to *Collaborative Design Patterns (CDPs)* [26]—a conceptual tool for thinking about typical learning situations. In particular, CDPs have been employed in the literature on computer-supported collaborative learning to characterize interaction patterns around the use of learning technologies such as handheld computers [27]. Collaboration patterns we observed with DIVER were emergent, and not behaviors that we prescribed or that were dictated by the software, yet they were

still comprised of the standard CDP elements (e.g., a problem, a context, a solution approach, etc.). Conceptualizing our observed patterns as CDPs is useful because it will allow us to make design suggestions for supporting specific kinds of user groups and to elicit patterns of behavior that have shown themselves to be particularly effective for such groups.

Here, we briefly describe the five most notable collaboration patterns we observed in our analyses of group activity in DIVER. Our data set encompasses numerous instances of each of these patterns, but we will describe each with a single example to elucidate its key features.

3.1 Collective Interpretation

Making interpretations of human behavior—particularly of learning—is a complex enterprise that benefits from multiple perspectives [28], [29]. The importance of collecting multiple interpretations is known instinctively by most, and so, when an individual sets out to conduct an analysis of a video event, he or she will often recruit the input of others who may bring novel insights based on their differences in knowledge and experience. The coordination of multiple perspectives, however, has long been a challenge for video analysts, especially in the days of VHS tapes, where a single individual had playback control and had to somehow “harvest” the insights of a room full of contributors [14]. The DIVER platform has affordances for sharing and comparing insights in a more coordinated manner, and we have observed that numerous groups use DIVER in an attempt to achieve interpretive consensus or to refine points of contention.

An example of this pattern of activity occurred within the context of a teacher credential program at a large east coast university. Students in a course on teaching and learning were provided with sample video of a science teacher working with her students to help them understand animal classification systems. The student-teachers who watched the video were tasked with describing how the instructional approach of the teacher in the video interacted with her students’ learning processes. By giving each student-teacher independent access to the focal event in the DIVER system, they were able to anchor their interpretations directly with video data rather than having to rely on their memory of what they had seen. A public record of these interpretations is maintained so that another contributor can offer a response hours or days later.

At one point in the science classroom video, the teacher asks one of her students about what they have learned from an animal classification activity. In the video, the student says: “It helped me realize what things have in common and what they don’t have in common.” One of the student-teachers (ST1) selected this clip from the video and moved it to the DIVER worksheet, which led to this exchange with another student-teacher (ST2):

ST1: *This student found the activity helpful in terms of understanding what an evolutionary biologist does. He seems to have gotten the big picture.*

ST2: *I don’t know if I agree that he has gotten the big picture. Though I admit that I could be thrown off by his affect, which is pretty flat, it seems like he is just saying the most basic things—we were putting out species in different categories, they [biologists]*

categorize certain species or animals, I found it helpful because it helped me realize what things have in common and what don’t. Is this the depth you were hoping for? How could we move him to the next level?

Importantly, the contribution of ST2 likely changed the interpretation of this student’s reflection, not just in the mind of ST1, but for all group participants who are now thinking about how the instructional intervention had failed to achieve deep student understanding. Had ST2 simply stated: “I disagree that the student has gotten the big picture” in an unanchored group discussion of the event, there would have been a greater chance that interpretive disagreement would have persisted. On the contrary, ST2 grounded her assertions in the actual classroom events, effectively shifting the analytical focus for subsequent interpretations.

We observe this pattern of collaboration in DIVER frequently. It typically emerges from groups who have the open-ended task of determining “what is going on” in a video event, ranging from field data collected by a research team to a funny video that someone posts on YouTube. The pattern often starts with one person taking a pass at making their own interpretation, and then others chime in with support or criticism. DIVER gives this discourse structure, such that an outside observer can fairly easily view a dive and ascertain the general consensus.

3.2 Distributed Design

There can be considerable value in collecting and considering video of the user experience in the design process [30], but the logistics of incorporating this video into a design team’s workflow can be tricky. We have encountered a number of well-intentioned design groups who collect hundreds of hours of video of users, only for this video to sit unwatched on a shelf of DVDs. Video can capture notable affordances or systematic failures in the user experience, but identifying these trends and communicating them to the rest of the team at critical points in the design process is an arduous, time-consuming task. It is not typically feasible for an entire design team to convene and spend hours watching videos of user testing. We have, however, observed that the members of a number of different design initiatives use DIVER as a tool for efficiently integrating video insights into their collaborative thinking and design reviews.

The distributed design pattern was observed in a small research team at a west coast university working on a prototype touch-screen interface for preschool children. The researchers were trying to create a tool that allowed children to construct original stories using stock video footage (e.g., baby animals in the wild). Access by the design team to children of this age was limited, but there was a real need for the team to understand how the children would react to this kind of interface and whether it was possible to scaffold their developing storytelling abilities with a novel technology. The design team arranged for a few pilot user sessions to be conducted at a local nursery school, but neither the lead researcher on the project nor the software engineer who was responsible for implementing interface changes were able to attend. The solution was for two graduate students to run the pilot sessions and



Fig. 3. DIVER worksheet showing an instance of distributed design.

immediately upload the session videos to DIVER so that the entire team could quickly formulate design modifications that could be implemented for the next iteration (Fig. 3).

In this instance, one team member took a first pass at segmenting and highlighting the important events in the video. This allowed the rest of the team to focus in on the moments that had potential design implications. In several instances, a team member would make a design recommendation based on something that they observed in the video, and one of the team members present at the session would respond with a clarification of what had occurred and possible alternative design schemes. The discourse in DIVER quickly transitioned from an investigation to a design plan that was adopted by the software engineer and implemented on a short timeline.

As a pragmatic matter, design teams in all areas of development will typically rely on data summaries and aggregations of user-test findings to inform the design process. While understandable, this practice can distance designers from the valuable insights to be derived from observing key moments in video recordings of users. A system for supporting asynchronous references of specific space-time segments of video—such as the one found in DIVER—seems to alleviate some of the coordinative barriers and permit the integration of user data with the design process. Recognizing this potential, a number of DIVER user groups found success generating new designs for learning and education through their discussions of video-recorded human interactions.

3.3 Performance Feedback

A challenge of giving people feedback on a performance, whether a class presentation or some display of artistic skill, is that the feedback offered is typically separated from the performance itself, such as a verbal evaluation given a week later or a review that has been written up in a newspaper. The effectiveness of the feedback in these cases relies on the memory of the performance both for the persons giving the review, as well as the person receiving the feedback. If, for example, a student does not recall misquoting a philosopher in their presentation for a law school course, they are not likely to be receptive to having this pointed out by their professor or classmates. Associating performance feedback

with actual segments or images from the performance has an intuitive utility, but the approach to implementing a feedback system that effectively makes these associations and integrates input from multiple individuals is less clear. Using DIVER as a means to deliver feedback on a video-recorded performance was one of the most common collaboration activities that we observed among our users. We were particularly impressed by the range of performance types (e.g., films produced by undergraduates, K-12 student-teaching) for which users attempted to use DIVER for conveying suggestions and making specific criticisms.

A good example of the performance feedback pattern is its use in a prominent US medical school where an effort was made to improve the manner by which students communicated important health information to patients. Medical students interning at a hospital were asked to record their consultation sessions with patients and submit these recordings to their mentors using DIVER. Experienced medical professionals offered these students constructive feedback on how they were communicating with their patients and provided suggestions for improvement. In some cases, there was also the opportunity for the students to respond to the feedback with questions or points of clarification. In one instance, a student uploaded her interview with a patient who came in with various ailments including abdominal pain. She asked numerous questions in her attempt to narrow down the underlying problem. In this dive, she did some self-evaluation—marking segments and making comments where she believes that she could have done something better. Additionally, her mentor (M) watched the video and marked his own segments upon which he based his evaluation. He marked one moment in the video in particular and made the following comment:

M: Good check here on the timing of her GI symptoms in the bigger picture—you are now making sure you know where these symptoms fit into the time course of this complicated history.

Note that the evaluator used the word “here” rather than having to describe the referent event in detail, as one would likely have to do if they were delivering an entirely written or verbal assessment. Targeted feedback of this kind should help to minimize misunderstandings or generalizations. People will sometimes overreact to negative feedback on their performance, concluding hastily that “she hated it” or “I can’t do anything right,” but with feedback that is linked to specific behaviors, there is a better chance that the evaluations will be received constructively. In a related vein, a recent paper described productive uses of DIVER for changing the paradigm of communication skills teaching in oncology to a more precise performance feedback system, rather than one principally based on observation [31].

Additionally, we observed users administering feedback as a group, such as when an entire class was asked to respond to a student presentation. In this case, the students were aware of and were able to coordinate their feedback with that offered by their classmates, leading to less redundancy. Users in such a design pattern also have the opportunity to mediate each other’s feedback, perhaps by giving support for another’s comments with additional ideas for improvement or by softening criticism that may be seen by some as overly harsh.

3.4 Distributed Data Coding

Some attempts at making interpretations of video recordings of human activity are more structured than those that we observed in the distributed interpretation pattern. In research settings in particular, the categories of activity that are of interest are often clear, while it is the identification of those categories and formal labeling of events in the data that must be negotiated. In their experience, coding video for the TIMSS study—a cross-cultural study of math and science classrooms—Stigler et al. [32] reflect on two lessons they learned about coding video data: 1) the videos must be segmented in meaningful ways to simplify and promote consistency in coding and 2) the construction and application of these codes requires input from multiple individuals with differing expertise. Both of these issues can be facilitated by a system like DIVER that structures video discourse around segments selected by any number of participants. While only a few user groups implemented a formal coding scheme in DIVER, their interaction style took on a distinct pattern that is useful to consider for thinking about possible applications and design needs.

One group of users that utilized DIVER to manage their distributed coding work was a team of seven researchers working as part of a large center dedicated to the scientific study of learning in both formal and informal environments. This team conducted over 20 interviews with families in the San Francisco Bay area on how uses of mathematics arise in their home life. The research group was interested in how families of diverse backgrounds organize their mathematical practices and to analyze how differences in social conditions and resources in the home support these practices. The videos from all of the interviews were uploaded into DIVER, but for all of the two-hour videos to have been logged and coded by the entire group would have been overly burdensome and ultimately counter-productive. Instead, the group reached consensus on a coding scheme by reviewing a subset of the interviews as a group, and then assigned two different team members to code each of the remaining videos. The codes that they agreed upon were labels for specific instances that they were interested in studying as part of their analysis, such as “gesture”—which denoted a family member had made some physical gesture to illustrate a mathematical concept, or “values”—which was used when someone made an expression of their family’s values when discussing a mathematical problem that arose in their home life. The consistent use of these codes in DIVER was particularly useful because the software allows users to search for keywords in the annotations of the video segments. The results of a search are frames *from multiple dives* where the word or code was used, meaning that this research team was able to quickly assemble all the instances of a particular code that had occurred across all the interviews, regardless of who had assigned the code. This facilitated the process of making generalizations across families and drawing conclusions that addressed their hypotheses.

While other tools for video analysis and coding exist (such as commercial systems ATLAS.ti, NVivo, and Studio-Code), most are not available online nor do they support multiple users in collaborative analyses. In its current state,

DIVER is probably not sufficient for largescale video coding—the research group described above, for example, supplemented their analysis with a FileMaker Pro database that held multiple data fields and detailed code specifications—but more sophisticated coding capabilities are not difficult to add and are currently in development. What is notable about the distributed coding pattern that we observed in existing DIVER users is that the key coding needs of segmenting and supporting multiple perspectives [32] were both supported and utilized.

3.5 Video-Based Prompting

Most of the collaboration patterns discussed so far have consisted of fairly open-ended interpretive work. Even the distributed coding pattern, though using a set of predefined categories, still required video segmentation and the reconciliation of coded events where there was disagreement. Some of the activity that we observed in DIVER, however, was far more constrained. We observed such a pattern most frequently in formal educational contexts such as classrooms where instructors had specific questions about a video event that they wanted their students to try and answer. For example, a film studies professor had a set of questions that he wished to pose to the students in his class about classic films like Godard’s *Breathless*. He used DIVER to distribute these questions because he wanted his questions *and his students’ answers* to be anchored by clips from the actual film. In some of the DIVER user groups, there was an instructor or facilitator that initiated a dive by capturing a clip from a larger video source and using the worksheet to pose a question about some aspect of the clip.

An especially interesting application of DIVER and a good example of the video-based prompting pattern was an undergraduate Japanese language course at a west coast university. The instructor of this course had collected a number of videos of informal spoken Japanese from various sources such as interviews and television programs. These videos demonstrated certain styles and forms of the language that the instructor wished her students to experience and reflect upon. These videos were uploaded into DIVER and the instructor created homework assignments where students would have to respond to four or five questions, each associated with a video segment selected by the instructor. In this class, the students were all able to see each other’s responses, which turned out to be a benefit in the eyes of the instructor because it stimulated the students to think deeply and try and make an original contribution to the analysis. Fig. 4 is a screenshot of the worksheet used for one of these homework assignments.

Instructors and facilitators of various kinds will frequently use visual aids such as video for encouraging critical thinking by their students. A Web services platform like DIVER allows students and participants to take these visual aids home with them, make independent interpretations, and then contribute their input in a structured forum. The important feature of this pattern is that any interpretation is necessarily associated with video evidence of the target phenomena. This practice has potential long-term benefits in that it teaches people to adequately support their arguments by making direct links to available data.



Fig. 4. A DIVER worksheet used in a Japanese language course. The instructor has posted questions for her students about informal language conventions used in the video.

4 MAPPING THE SPACE OF COLLABORATIVE VIDEO ANALYSIS TASKS

There are numerous learning situations that can be enhanced with video analysis, resulting in different patterns of collaboration, each with different needs in terms of interface supports and structure. DIVER's open-ended platform allowed many of these patterns to emerge organically, but it was clear that several of these collaborative practices could be enhanced with additional features and capabilities. While DIVER has taken modest steps forward on this front—adding coding, transcription, and clip trimming functionality, for example—we recognize that there is still much progress to be made to fully exploit the learning potential of digital video for researchers and for education.

In order for the community of learning technology researchers and developers to adequately address this design problem, we have attempted to map out the space of collaborative video practices. In extracting the five collaboration patterns from our data set, we recognized a few salient dimensions that we can abstract from these practices for defining this space. These dimensions capture all of the practices we observed, but more importantly, they suggest a number of other practices that we did not observe. Perhaps due to limitations in DIVER or video technologies in general, or simply incidental to the needs of the DIVER user community to date, there were several possible and promising collaboration activities that have not yet manifested themselves in mainstream video practices. By specifying the features of this space in its entirety, our hope is that we can more fully address the design needs of existing practices as well as cultivate fledgling practices that could have a significant impact on the culture of learning technology. The three dimensions that define this space are *the style of discourse*, *relationship to the source material*, and *target outcome*.

4.1 Discourse Style

In discussing and interpreting the events within a video, the needs of some groups are best met by employing a more informal structure with fewer constraints on participation.

Exploratory analyses or discourse around video for purely social purposes are likely to adopt this more conversational style. Other groups, however, such as in courses, have specific aims for their collaboration and participatory roles that must be maintained during the course of discussing the video. These groups will adopt a more formal discourse style that may involve prompting participants for desired input or limiting contributions to specific times or locations within the video.

4.2 Relationship to Source Material

Digital video technologies have made tremendous advances over the last decade, particularly in the capacity for capturing, uploading, and sharing video at rapid speeds and with relatively little effort. The ubiquity and ease of transmission of digital video has changed the traditional relationships one has with the video they watch for recreation or utilize in a professional capacity. On one end of this dimension is an insider relationship, meaning that the video used for the collaborative activity is video with which the group has strong familiarity. It could be video that someone in the group recorded or video in which the group members themselves are featured. With insider videos, the group often has some degree of control over how the video was recorded and for what purpose. If a group has an outsider relationship with a source video, they typically have less control over factors such as editing style and camera orientation. These are videos that may have been obtained from a secondary source or were selected from an archive collection. It is less likely that groups discussing outsider videos will possess background information or be able to fill in comprehension gaps stemming from contextual features and events that were not recorded.

4.3 Target Outcome

Any group of individuals that sets out to explore a video record asks themselves: "What do we want to get out of this?" As should be apparent from our review of collaborative video activities in DIVER, there are numerous possible answers to this question. We have identified four general types of outcomes that, while not exhaustive or mutually exclusive, describe the bulk of possible collaboration objectives: *design*, *synthesis/pattern finding*, *evaluation*, and *analysis/interpretation*. Groups are working toward design outcomes when they discuss video with the aim of conceiving or improving upon some product, process, or organizational scheme. Synthesis or pattern-finding outcomes come from attempts to reach consensus on "the big picture" and recognize important trends and commonalities. Evaluation outcomes are critiques of video products and recorded events with the corresponding aim of either delivering feedback to a specific group or for demonstrating critical competence (as in a K-12 teacher performance assessment). Unlike synthesis outcomes, groups with an analysis or interpretation objective are trying to "break things down" and typically are attempting to understand what is happening behind the scenes or in the minds of the actors that is causing the events seen in the video.

Explicating these dimensions results in a matrix of collaboration activities represented in Fig. 5. We caution that the states of each dimension that we have identified

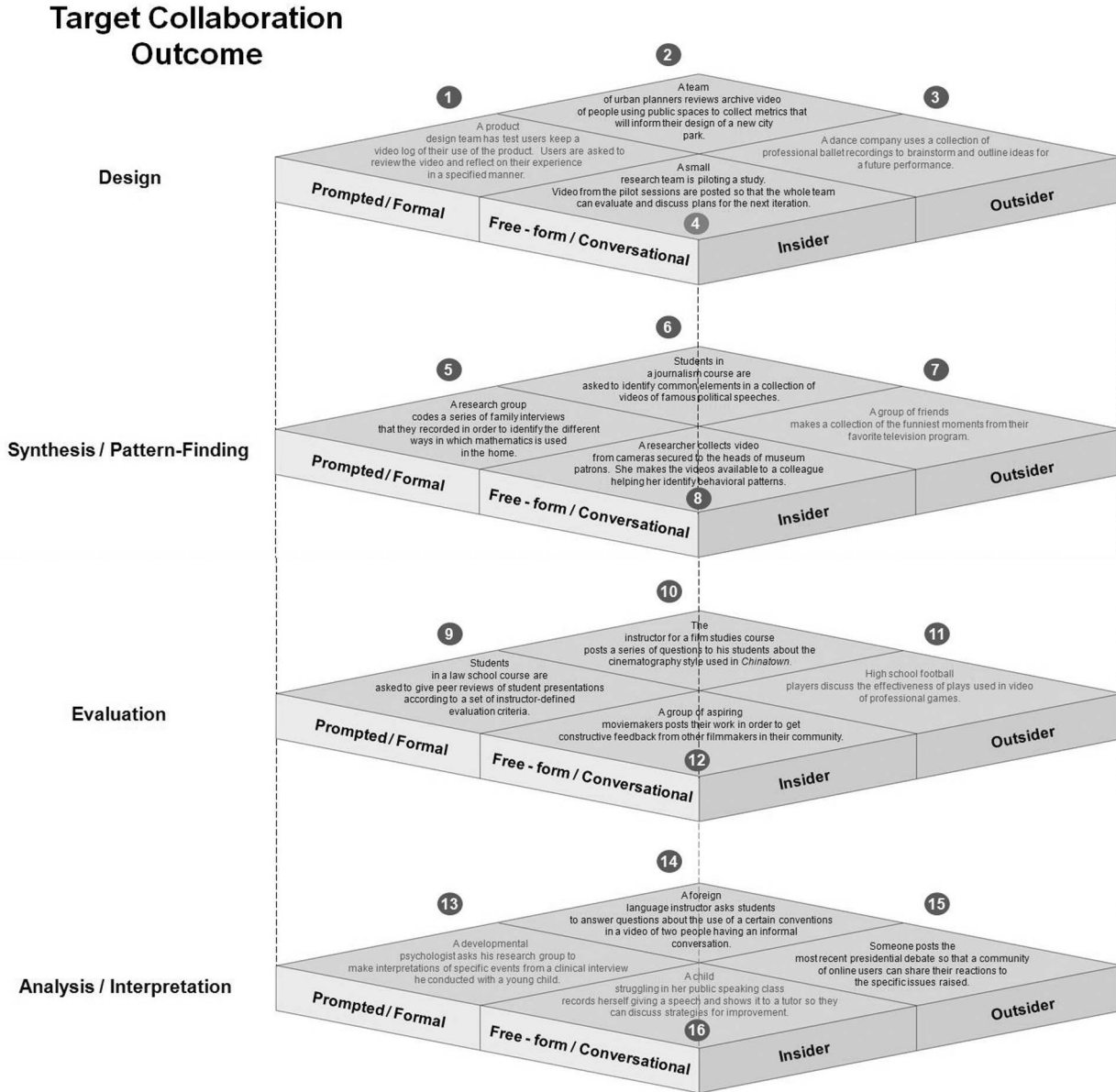


Fig. 5. A representation of possible design patterns based on three dimensions of collaborative activity: target outcome, discourse style, and relationship to the source material. Cells with black print indicate collaboration patterns that we have observed on the DIVER platform. Cells with grey print describe hypothetical scenarios that have not yet been observed but are plausible given the appropriate user group and software environment.

here do not have “hard” boundaries—it is perfectly feasible that a collaboration activity around video could straddle the line between a formal and informal discourse style, for example. However, we believe that this representation provides a useful depiction of the range of possibilities and the various needs of groups that fall within this space. To clarify these possibilities and needs even further, we have offered examples of activities that embody the defining characteristics of each cell. Some of these examples are actual activities that we have observed in DIVER and some are hypothetical activities that share the same characteristics of those that we have observed. Other examples in this representation (in grey typeface) are hypothetical activities with characteristics that we have not yet observed directly in DIVER but are possible presuming that the right

configuration of situational variables exists. All the examples offered in Fig. 5 are situated within a particular user group (teenagers, industry professionals, etc.), but again, we reiterate our conviction that these activity patterns could be implemented in any number of educational and research settings.

These activities should be of particular interest to designers and educators because they present powerful learning opportunities with digital video that simply have not been realized with current toolsets. In the meantime, there is still a great deal of work that can be done to address the activity patterns that we have observed, whether they are minor modifications to a software platform like DIVER or the development of an entirely new technology that targets a specific subset of the space we have defined here.

5 DESIGNING FOR VIDEO-BASED COLLABORATIONS

When a group or a team is planning to embark on a project using video as an “anchor” [33] for their collaborative activity, they need to have in mind the target outcomes of their work, and as we have seen, such envisioned outcomes may involve design, synthesis/pattern finding, evaluation, and analysis/interpretation. Now that we have a handle on not only actual design patterns from uses of the DIVER video platform, but a three-dimensional heuristic matrix for generating possible design patterns which vary in terms of target outcome, relationship to video sources (insider/outsider), and discourse style (formal/informal), we can visit the question of what forms of supportive structures might serve as useful design scaffolding for the activities of video-based collaborative groups.

5.1 Designs for Target Outcomes

We can envision a number of support structures for groups whose objective is to produce one of the four outcomes we have identified in the matrix. For groups that are aiming at *design*, it would be useful to have platform capabilities that support the processes of iteration and revision. Videos that capture product development at different stages, for example, could be tagged as such in a video database. Support for design argumentation [34] could be integrated, linking to video evidence. A collaboration platform for supporting design could also include features for tracking progress, such as milestone completion markers that are anchored by video of a successful user-test. Additionally, video could be accompanied by a “design canvas” that allows users to make free-form sketches or models that are shared with the group. Rather than the worksheet in DIVER, for example, users could have a sketch space where they could manipulate images captured on the video to convey new ideas.

For *synthesis and pattern-finding* objectives, there is a need for tools that aid in collecting and building relationships between instances of a focal interest. With rapid advances underway in computer vision technology [35], it is becoming feasible that recognition algorithms could be used to automatically identify key objects, faces, head poses, and events in video (e.g., hand-raising in a classroom discourse, uses of math manipulatives, etc.) and generate tagging metadata for videorecords. Modules for a video research platform that could reliably flag such moments of analytic content automatically and make them readily available for subsequent analyses would save immense time and effort.

Evaluation outcomes could be aided with tools that minimize redundancy and make the provision of feedback more efficient. Rather than free-form text input, it may be appropriate to provide a template for a rating system or checklist schemas. To avoid the recurrence of the same comment, users could be enabled with a way to show support for an existing comment or poll functionality—a button that says something to the effect of “I agree with that.” Another need that we have observed in groups doing evaluation is the integration of documents and other materials associated with a performance. Someone being evaluated as part of the teacher credentialing process, for example, may be asked to submit lesson plans and

examples of student work in addition to video of their lesson. A platform that allows evaluators to dive into and annotate these supplemental documents and connect them to the video event could provide for a more comprehensive workflow for assessment processes.

In *analysis* activities, participants are typically looking for ways to contribute the most insightful information in the most efficient manner possible. While text can be a good way of sharing interpretations, there are some scenarios where voice annotations would be more effective at conveying a nuanced perspective [18]. There are also some situations where an analysis is best supported with an existing document or representation. In this case, having the ability to simply link a video event to a URL or to embed an object in the analysis would be advantageous. Finally, some analyses require one to look at an event from multiple vantage points. Having the ability to easily synchronize multiple-video sources and view simultaneous playback is a feature requested by a number of our user groups. Researchers of human-computer interaction, for example, may want to study how someone uses a new car prototype using video streams of the front-view, the rear-view, and the driver’s face and posture.

5.2 Designs for Relationships to Video Sources

Both insider and outsider relationships with source video would likely benefit from different metadata capabilities. For *insiders* who have shot their own video and wish to share it with a small group or broadcast it to the whole world, it is important to be able to control the information, or metadata, that people have about that video. Where was the video recorded? Who is featured in the video? For what purpose was it recorded? This is also a potential opportunity to specify how the video can and should be used (e.g., granting *Creative Commons* licensure). This is true not only for the video itself, but for the analysis that one performs on the video. These analyses can themselves be works of intellect and artistic expression, and so attribution of this work is important.

For *outsiders*, it would be highly valuable to have access to any metadata that was provided for video a group is using. Geographic data about the video, for example, could facilitate mashups of video and mapping tools such as Google Earth. If a group was working with a large archive of video, it could be useful to have a platform that uses timestamp data to organize and create visualizations based on when the video was recorded.

5.3 Designs for Discourse Style

If the desired type of *formal discourse* for a certain group is known in advance, it would be possible for a video collaboration platform to structure this type of discourse using templates or other interface constraints to “script” activities [36]. Some forms of collaboration have explicit rules for how these activities should be conducted, and there is potential for the interface to assist in regulating this activity as a form of distributed intelligence [37], [38]. Besides templates, this effect can be achieved by assigning participants different roles with associated permissions and capabilities, or there can be features that support divisions

of labor by directing participants to work on different parts of the video task.

Groups that desire a more *informal or conversational discourse style* will likely desire fewer constraints and more flexibility for communicating and constructing new insights. This may include features that allow for more social interaction such as chat or networking capabilities (e.g., you may want to connect with someone if you know that they have extensive experience doing a certain type of video analysis). It may also include tools for doing more free-form interpretive work, such as “build-as-you-go” coding schemes. These types of capabilities may allow for the emergence of novel designs and interpretations that would not develop in a more constrained setting.

6 CONCLUSION

In this paper, we have articulated a vision for the generative promise of video research and education platforms for supporting the work practices of collaborative groups. In particular, the DIVER video platform embodies a new kind of communication infrastructure for *video conversations* by providing persistent and searchable records of video pointing activities by participants to specific time and space moments in video so as to develop “common ground” in technology-mediated conversations among distributed teams. Rather than speculate about the opportunities for using the DIVER platform, we empirically examined how roughly 200 globally distributed groups appropriated the DIVER technology to serve their needs. We found five dominant *collaborative design patterns*: collective interpretation, distributed design, performance feedback, distributed data coding, and video-based prompting. Abstracting from the features of these collaboration patterns, we were able to identify a set of three dimensions that we argue provide a tripartite design space for video collaboration groups: *discourse style* (formal/informal), *relationship to video source* (insider/outsider), and *target outcome* (design, synthesis, evaluation, and analysis). These dimensions were used heuristically to articulate a three-dimensional matrix for conceptualizing video collaborative groups, and then used to spawn concepts for dimension couplings unrealized in DIVER collaborative groups to date, and to recommend new socio-technical designs for better serving the design space of these groups. We invite the learning technologies community to refine and advance our conceptualizations of collaboration design patterns for video platform uses in research and education and the creation of systems that support the important needs for video conversations in the work practices of educators and researchers.

ACKNOWLEDGMENTS

The authors give special thanks to DIVER software engineer Joe Rosen for his exceptional contributions to every aspect of the DIVER Project since 2002, and Michael Mills, Kenneth Dauber, and Eric Hoffert for early DIVER design innovations. DIVER, WebDIVER, and Guided Noticing are trademarks of Stanford University for DIVER software and

services with patents awarded and pending. The authors are grateful for support for The DIVER Project in grants from the US National Science Foundation (#0216334, #0234456, #0326497, #0354453) and the Hewlett Foundation.

REFERENCES

- [1] P. Saettler, *The Evolution of American Educational Technology*. Information Age Publishing, 2004.
- [2] *Video Research in the Learning Sciences*, R. Goldman, R.D. Pea, B. Barron, and S. Derry, eds. Lawrence Erlbaum Associates, 2007.
- [3] T. Chambel, C. Zahn, and M. Finke, “Hypertext and Cognition: Designing Video-Based Hypermedia for Individual Learning and Collaborative Knowledge Building,” *Cognitively Informed Systems: Utilizing Practical Approaches to Enrich Information Presentation and Transfer*, E.M. Alkhalifa, ed., pp. 26-49, Idea Group, 2006.
- [4] R. Goldman, “Video Perspectivity Meets Wild and Crazy Teens: A Design Ethnography,” *Cambridge J. Education*, vol. 34, no. 2, pp. 157-178, June 2004.
- [5] D. Schwartz and K. Hartman, “It’s Not Television Anymore: Designing Digital Video for Learning and Assessment,” *Video Research in the Learning Sciences*, R. Goldman, R. Pea, B. Barron, and S.J. Derry, eds., pp. 335-348, Erlbaum, 2007.
- [6] C. Zahn, R. Pea, F.W. Hesse, M. Mills, M. Finke, and J. Rosen, “Advanced Digital Video Technologies to Support Collaborative Learning in School Education and Beyond,” *Computer Supported Collaborative Learning 2005: The Next 10 Years*, T. Koschmann, D. Suthers, and T-W. Chan, eds., pp. 737-742, Erlbaum, June 2005.
- [7] R. Pea and E. Hoffert, “Video Workflow in the Learning Sciences: Prospects of Emerging Technologies for Augmenting Work Practices,” *Video Research in the Learning Sciences*, R. Goldman, R. Pea, B. Barron, and S.J. Derry, eds., pp. 427-460, Erlbaum, 2007.
- [8] B. MacWhinney, “A Transcript-Video Database for Collaborative Commentary in the Learning Sciences,” *Video Research in the Learning Sciences*, R. Goldman, R. Pea, B. Barron, and S.J. Derry, eds., pp. 537-546, Erlbaum, 2007.
- [9] R. Goldman, “Orion, an Online Collaborative Digital Video Data Analysis Tool: Changing Our Perspectives as an Interpretive Community,” *Video Research in the Learning Sciences*, R. Goldman, R. Pea, B. Barron, and S.J. Derry, eds., pp. 507-520, Erlbaum, 2007.
- [10] R.M. Baecker, D. Fono, and P. Wolf, “Towards a Video Collaboratory,” *Video Research in the Learning Sciences*, R. Goldman, R. Pea, B. Barron, and S.J. Derry, eds., pp. 461-478, Erlbaum, 2007.
- [11] H. Jenkins, K. Clinton, R. Purushotma, A.J. Robinson, and M. Weigel, *Confronting the Challenges of Participatory Culture: Media Education for the 21st Century*. The MacArthur Foundation, <http://tinyurl.com/y7w3c9>, 2006.
- [12] *How People Learn: Brain, Mind, Experience, and School (Expanded Edition)*, J.D. Bransford, A.L. Brown, and R.R. Cocking, eds. National Academy Press, 2000.
- [13] J. Bransford, N. Vye, R. Stevens, P. Kuhl, D. Schwartz, P. Bell, A. Meltzoff, B. Barron, R. Pea, B. Reeves, J. Roschelle, and N. Sabelli, “Learning Theories and Education: Toward a Decade of Synergy,” *Handbook of Educational Psychology*, second ed., P. Alexander and P. Winne, eds., pp. 209-244, Erlbaum, 2006.
- [14] B. Jordan and A. Henderson, “Interaction Analysis: Foundations and Practice,” *J. Learning Sciences*, vol. 4, no. 1, pp. 39-103, Jan. 1995.
- [15] S. Derry, R. Pea, B. Barron, R.A. Engle, F. Erickson, R. Goldman, R. Hall, J. Lemke, M.G. Sherin, B.L. Sherin, and T. Koschmann, “Guidelines for Conducting Video Research in the Learning Sciences,” *J. Learning Sciences*, 2009 (in press).
- [16] R. Pea, “Video-as-Data and Digital Video Manipulation Techniques for Transforming Learning Sciences Research, Education, and Other Cultural Practices,” *The International Handbook of Virtual Learning Environments*, J. Weiss, J. Nolan, J. Hunsinger, and P. Trifonas, eds., pp. 1321-1393, Springer, 2006.
- [17] S.E. Page, *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. Princeton Univ. Press, 2007.
- [18] R. Stevens, “Capturing Ideas in Digital Things: A New Twist on the Old Problem of Inert Knowledge,” *Video Research in the Learning Sciences*, R. Goldman, R. Pea, B. Barron, and S.J. Derry, eds., pp. 547-563, Erlbaum, 2007.
- [19] R. Pea, M. Mills, J. Rosen, K. Dauber, W. Effelsberg, and E. Hoffert, “The DIVER™ Project: Interactive Digital Video Repurposing,” *IEEE Multimedia*, vol. 11, pp. 54-61, 2004.

- [20] R. Pea, R. Lindgren, and J. Rosen, "Cognitive Technologies for Establishing, Sharing and Comparing Perspectives on Video over Computer Networks," *Social Science Information*, vol. 47, no. 3, pp. 355-372, Sept. 2008.
- [21] M. Lampert and J. Hawkins, "New Technologies for the Study of Teaching," Report to the US National Science Foundation from a workshop held 9-11 June 1998, NSF Grant #REC-9802685, June 1998.
- [22] B. MacWhinney and C. Snow, "Multimodal Studies of Instructional Discourse," Report to the US National Science Foundation, <http://www.nsf.gov/sbe/tcw/events991203w>, 1999
- [23] R.D. Pea and K. Hay, "CILT Workshop on Digital Video Inquiry in Learning and Education," Report to the US National Science Foundation based on NSF #0124012, Stanford Center for Innovations in Learning, 2003.
- [24] H.H. Clark, *Using Language*. Cambridge Univ. Press, 1996.
- [25] W.V.O. Quine, "The Inscrutability of Reference," *Semantics: An Interdisciplinary Reader*, D. Steinberg and L. Jacobovits, eds., pp. 142-154, Cambridge Univ. Press, 1971.
- [26] C. DiGiano, L. Yarnall, C. Patton, J. Roschelle, D. Tatar, and M. Manley, "Conceptual Tools for Planning for the Wireless Classroom," *J. Computer Assisted Learning*, vol. 19, no. 3, pp. 284-297, Sept. 2003.
- [27] T. White, "Code Talk: Student Discourse and Participation with Networked Handhelds," *Computer-Supported Collaborative Learning*, vol. 1, no. 3, pp. 359-382, Aug. 2006.
- [28] S. Wilson, "The Use of Ethnographic Techniques in Educational Research," *Rev. Educational Research*, vol. 47, no. 2, pp. 245-265, Spring 1977.
- [29] R. Goldman-Segall, *Points of Viewing Children's Thinking: A Digital Ethnographer's Journey*. Erlbaum, 1998.
- [30] J. Buur and K. Bagger, "Replacing Usability Testing with User Dialogue," *Comm. ACM*, vol. 42, no. 5, pp. 63-66, May 1999.
- [31] A.L. Back, R.M. Arnold, W.F. Baile, J.A. Tulsy, G. Barley, R.D. Pea, and K.A. Fryer-Edwards, "Faculty Development to Change the Paradigm of Communication Skills Teaching in Oncology," *J. Clinical Oncology, The Art of Oncology*, Mar. 2009.
- [32] J.W. Stigler, R. Gallimore, and J. Hiebert, "Using Video Surveys to Compare Classrooms and Teaching Across Cultures: Examples and Lessons from the TIMSS Video Studies," *Educational Psychologist*, vol. 35, no. 2, pp. 87-100, June 2000.
- [33] J.D. Bransford, R.D. Sherwood, T.S. Hasselbring, C.K. Kinzer, and S.M. Williams, "Anchored Instruction: Why We Need It and How Technology Can Help," *Cognition, Education and Multimedia*, D. Nix and R. Sprio, eds., pp. 163-205, Erlbaum, 1990.
- [34] S. Buckingham Shum, "Design Argumentation as Design Rationale," *The Encyclopedia of Computer Science and Technology*, vol. 35, Supp. 20, pp. 95-128, Marcel Dekker Inc., 1996.
- [35] D.G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int'l J. Computer Vision*, vol. 60, no. 2, pp. 1573-1405, 2004.
- [36] P. Dillenbourg and P. Tchounikine, "Flexibility in Macro-Scripts for Computer-Supported Collaborative Learning," *J. Computer Assisted Learning*, vol. 23, no. 1, pp. 1-13, 2007.
- [37] R.D. Pea, "Practices of Distributed Intelligence and Designs for Education," *Distributed Cognitions*, G. Salomon, ed., pp. 47-87, Cambridge Univ. Press, 1993.
- [38] J. Hollan, E. Hutchins, and D. Kirsch, "Distributed Cognition: Toward a New Foundation for Human-Computer Interaction Research," *ACM Trans. Computer-Human Interaction (TOCHI '00)*, vol. 7, no. 2, pp. 174-196, 2000.



Roy Pea is a professor of learning sciences at Stanford University and the H-STAR Institute codirector (<http://hstar.stanford.edu>). He has published extensively on learning and education fostered by advanced technologies, including scientific visualization, online communities, digital video collaboratories, and mobile learning. He is a co-PI the LIFE Center (<http://life-slc.org>), funded by the US National Science Foundation as one of several national Science of Learning Centers, and was coauthor of the 2000 National Academy Press volume *How People Learn*. He is a fellow of the National Academy of Education, the American Psychological Society, and the American Educational Research Association. He is a member of the IEEE and the IEEE Computer Society.



Robb Lindgren received the BS degree in computer science from Northwestern University in 2000 and is currently a doctoral candidate in the Learning Sciences and Technology Design program at Stanford University. He is currently researching learning and interactive media technologies at the LIFE Center while completing his dissertation on perspective-based learning in virtual worlds. He is a student member of the IEEE and the IEEE Computer Society.