CHI 2013 Workshop: Evaluation Methods for Creativity Support Environments

April 28, 2013

Paris, France

Creativity refers to human processes that underpin sublime forms of expression and fuel innovation. While a 2005 NSF workshop saw development and evaluation of creativity support tools as a new field, we take the position that researchers are developing sophisticated methods, which have progressed well beyond infancy. We expand the scope with 'environments', a superset of 'tools'.

Creativity support environments (CSEs) span and integrate diverse domains and types of systems, including software, hardware, instrumented spaces, networked topologies, and mobile devices. CSEs may involve temporal-spatial dimensions of collaborative work, requiring evaluation methods that address synchronous, asynchronous, co-located, and distributed interaction.

We seek to gather the community of researchers developing and evaluating CSEs, to share approaches, engage in dialogue, and develop best practices. We seek papers and presentations that develop in-depth methods for CSE evaluation. Methods must be explained with sufficient clarity and detail that others can apply them. They should be grounded by showing how they have been applied in the study of particular CSEs. Authors should motivate the types of CSEs for which a particular evaluation method is well-suited.

The workshop will coalesce the community involved in developing these methods, and set the table, inspiring discussion and debate about the value of particular methods in types of situated contexts. The expected outcome is not a single prescription, but a landscape of routes, an ontology of methodologies with consideration to how they map to creative activities, and an emerging consensus on the range of expectations for rigorous evaluation of CSE research.

We envision a lasting impact from this workshop, addressing shortcomings derived from lack of reuse of CSE evaluation methods in different contexts beyond the initial use. The workshop will create a plan for building an open repository of CSE evaluation methodologies, testing data, and descriptions of situated contexts where methods were applied. The discussed methodologies will form a basis for the repository.

Organizers

Program Chairs

Andruid Kerne, Interface Ecology Lab, Texas A&M University, USA

Celine Latulipe, HCILab, University of North Carolina at Charlotte, USA

Steven M. Drucker, Microsoft Research, USA

Linda Candy, University of Technology, Sydney, Australia

Kristina Höök, Swedish Institute of Computer Science, Kista, Sweden

Erin Carroll, HCILab, University of North Carolina at Charlotte, USA

Andrew M. Webb, Interface Ecology Lab, Texas A&M University, USA

Program Committee

Steve Benford, University of Nottingham

Win Burleson, Arizona State University

Sally Jo Cunningham, University of Waikato

Ellen Do, Georgia Institute of Technology

Krzysztof Gajos, Harvard University

Ashok Goel, Georgia Institute of Technology

Mary Lou Maher, University of North Carolina at Charlotte

Ali Mazalek, Georgia Institute of Technology

Frank Nack, University of Amsterdam

Kumiyo Nakakoji, Software Research Associates Inc., Japan

Steven M. Smith, Texas A&M University

Krzysztof Gajos, Harvard University

Agenda

9:00 - 9:30: Introduce yourself

9:30 - 10:00: Workshop Instigators' Methodologies

Quantitative Evaluation Approaches: Self-Report and Physiological Responses
Erin Carroll and Celine Latulipe

Evaluating Information-based Ideation
Andrew M. Webb, Andruid Kerne, Rhema Linder, Steven M. Smith, Nic Lupfer and Yin Qu

10:00-10:45:Paper Presentations I (10+5 each)

MAIA: A Methodology for Assessing Imagination in Action
Sharon Lynn Chu and Francis Quek

Evaluating Creativity Support in Co-Design Workshops
Graham Dove and Sara Jones

Supporting Creativity in Performance Lectures
John V H Bonner and David Peebles

10:45 - 11:00: Poster Session

Intelligent Creativity Support
Winslow Burleson and Kasia Muldner

Evaluation Criteria for Safe Improvisation in EMS Technologies
Steven Haynes, John M. Carroll and David R. Mudgett

Creativity Support in Authoring and Backtracking
Brad Myers, Stephen Oney, Youngseok Yoon and Joel Brandt

A Tiered Evaluation Framework for Reality-Based Creativity Support Environments
Orit Shaer and Consuelo Valdes

Understanding Movement: the Design and Evaluation of Lighting and Wearable Technology in Dance Performances
Renata Sheppard and David A. Shamma

11:00 - 11:30: Coffee Break + Poster Session

11:30 - 12:15: Small Group Discussion

12:15 - 14:00: Lunch

14:00 - 15:00: CSE Evaluation Collaboration Speed Dating

Bring ideas for an interesting CSE you would like to create and an evaluation method (for that CSE or something else). During this session, pairs of participants will quickly discuss one's CSE and how to employ the other's evaluate method with it.

15:00 - 15:30: Coffee Break + Poster Session

15:30 - 16:30: Paper Presentations II (10+5 each)

Searching to Measure the Novelty of Collected Ideas
Rhema Linder, Andrew M. Webb and Andruid Kerne

The Observable Creative Process and Reflection-on-action
Vincent Akkermans, Geraint A. Wiggins and Mark B. Sandler

Techniques for Evaluating Novice-Oriented Creativity Support Tools
Nicholas Davis, Alexander Zook, Friedrich Kirschner, Mark Riedl and Michael Nitsche.

Design Creativity: Using Pareto Analysis and Genetic Algorithms to Generate and Evaluate Design Alternatives
Erin Bradner and Mark Davis

16:30 - 17:00: Small Group Discussion

17:00 - 18:00: Summative Discussion & Repository Planning

MAIA: A Methodology for Assessing Imagination in Action

Sharon Lynn Chu
Center for Human-Computer Interaction
Virginia Tech
sharilyn@vt.edu

Francis Quek
Center for Human-Computer Interaction
Virginia Tech
quek@vt.edu

ABSTRACT

We propose MAIA, a methodology to evaluate embodied storytelling systems aimed at supporting the child's creativity. Our approach rests on three key tenets: the study of process instead of objects, imagination as the basis of creativity, and creativity as recombination. MAIA uses micro-level analyses of enactment videos, the child's drawings and interview transcripts to generate a 'broader imagination' score that can be used for both quantitative and qualitative comparisons.

Author Keywords

Evaluation, Imagination, Children, Storytelling

ACM Classification Keywords

H.5.2. Information interfaces and presentation (HCI): Evaluation/methodology

INTRODUCTION

This paper proposes MAIA, a method to evaluate Creativity Support Environments (CSEs) through the study of the imagination at its point of occurrence. We address embodied systems that employ meaningful movements and gestures (or enactment) for interaction. We present our method in the context of technology-mediated children's storytelling. Many such storytelling systems involve the use of physical activity for interaction (e.g., Storymat [1], POGO [2], StoryToy [3] and StoryRoom [4]), and increasingly more so with the advent of innovatons like the *Wii* and *Kinect*. We first describe three key tenets that make up the philosophy behind our proposed approach and a brief review of previously used evaluation methods. We then explain our approach proper and present a case application, before concluding.

PHILOSOPHY OF MAIA

Tenet I: The Study of Process

"To encompass in research the process of a given thing's development in all its phases and changes…fundamentally means to discover its nature, its essence, for 'it is only in movement that a body shows what it is'" [5]. This quote from Lev Vygotsky illustrates the importance of studying processes instead of outcomes. We adopt this philosophy to look at creativity in the context of storytelling activities as it develops through the process of imagination.

We consider a CSE to be any system, artifact or object that seeks to facilitate the creative process that one engages in for a particular activity. A CSE can be evaluated in several ways – in terms of: a) How it supports the **person** to be more creative; b) How it enables the generation of creative **products**; c) How well it mediates the creative **process**; or d) Whether the technology furnishes a more effective **environment** for creativity.

A number of existing methods could be used for approaches a), b) and d). We found few methods adopting the approach c). Most that adopt **process** evaluation approaches employ a general mixed-methods framework. We found, however, that these lack specificity for children and do not provide exact guidance for our studies with embodied systems. Our approach seeks to evaluate the dynamic process (rather than the static products) of creative activity.

Tenet II: Imagination as the Basis of Creativity

Vygotsky places imagination at the genesis of the creative process stating: "the entire world of human culture … is the product of human imagination and of creation based on this imagination" [6]. Imagination, the ability to 'image' or 'see' distal or completely constructed worlds and possibilities, plays a key role: "every invention, whether large or small, before being implemented, embodied in reality, was held together by the imagination alone" [6]. Pelaprat & Cole [7] propose that imagination and creativity have a 'cyclical' relationship that is mediated by experience and knowledge. Experience shapes imagination, and imagination contributes to creative activity. If the output of creative activity is "perceived as new the products of imagination become creative when they enter the cultural world of interaction." [7]. Given its importance, we focus our approach on evaluating imagination, the root of creative activity.

Tenet III: Imagination and Creativity as Recombination

We observe that although they provide differing definitions, many researchers agree that creativity consists of a recombination of pieces from previously internalized experiences. Finke et al. [8] list the process of 'conceptual association and recombination' as the two important properties of creative cognition. Similarly, Boden [9] refers to 'combinatorial creativity', i.e. utilizing rules in a conceptual search space in new ways "to come up with new combinations. Fauconnier and Turner's [10] concept of 'conceptual blending' holds the same idea of combining domains. Vygotsky, in turn, calls creative activity, based on the ability of our brain to combine elements, imagination or fantasy [6]. Our approach embraces this idea of creativity as one engaging in a combinatorial process of 'idea bits'.

EVALUATING IMAGINATION

Imagination has been assessed in many different ways depending on the requirements of the study in question. The issue of a general assessment measure of imagination is complicated by the various ways in which the concept has been understood (e.g. it has previously been equated with memory, imagery, fantasy, invention or creativity). Some of the common measures that have been used include the numerous types of inkblot tests (Rybakoff, Whipple, Rorschach, etc.), textual measures (sentence building, story creation based around certain words, descriptions of imaginary animals, compositions, theme writing), studies of dreams and fantasy [11], or various scales depending on the definition adopted. Most of these measures are either not suitable for children, or do not measure imagination in-situ. From our first tenet, we are particularly interested in studying imagination in the process of its evocation. More importantly however, we could not find a method for analyzing imagination from gestures or enactment in the literature.

Loke et al. [12], Andrienko et al. [13] and the Laban Movement Analysis framework provide some indication as to how to analyze movement qualitatively but does not relate it to imagination in any way. Nemirovsky et al. [14] relate gestures to imagination using a method of analysis from psycholinguistics called microethnography, a "collection of techniques that focus on moment-to-moment bodily and situated activity". It is not always desirable, or even possible, to elicit co-produced speech from children enacting during the use of a system. We believe thus that a method of enactment analysis to elucidate the child's in-situ imagination is a substantial contribution.

MAIA: ASSESSING IMAGINATION IN ACTION

We devised a measure of imagination that we call 'broader imagination'. Broader imagination can be defined as including *any form of extension and association made beyond (visual, auditory, tangible, etc.) presented materials for the task at hand*. These extensions and associations can vary on <u>richness</u>, <u>typicality</u> with regard to a given situation and <u>consistency</u> over time. We use a multi-track approach to measure situated broader imagination from methods chosen to minimize the barrier of expression for children:

a. **Story enactments:** The child is given the task of telling a story with the CSE under study;
b. **Scene drawings:** After enacting, the child is asked to draw a scene that she enacted;
c. **Oral recall interviews:** At the end of the session, the child is engaged in a semi-structured interview about her enactments.

Data analysis in MAIA encompasses levels of micro-analysis of video and audio recordings, and collation of drawings. The video analysis serves to evaluate the richness of the enactment. We discern two types of analysis that moves from the objective level to the interpretive level:

1. **Generating Micro-Actions (MAs):** MAs represent the objective actions that the child performs in an enactment and consist of any action distinguishable as a unit.
2. **Generating Vignettes (Vigs):** Vigs represent the semantic interpretation of the set of micro-actions in the enactment: the mini-stories that the child is trying to tell. The interpretation of MAs into Vigs takes into account body postures, facial expressions, gaze, pace, etc.

We use analyses from the two other sources, interviews and drawings, to consolidate conclusions from the enactment analysis. Both serve as tests of richness and consistency of the imagination. The interview is used to probe about thoughts of the child during her enactment. Transcribed with timecodes, the interview is coded for four dimensions: the child's stated goal in the enactment (**goal**); the child's operationalization of the goal (**schema**); extra details that the child imagined (**extension**); and how consistent the child was during the interview in terms of intent, action and recall with regard to the enactment (**consistency**).

The drawings are coded for three dimensions: characters, if any, in terms of their suggested action (**character**); the scene or environment and any other elements in it (**scene**); and how consistent the child was in her drawing with regard to the enactment and the interview (**consistency**).

Finally, an 'overall broader imagination score' (OBIS) is given for each enactment, based on a gestalt view built from all of the child's enactment's micro-actions and vignettes, interview analysis and drawing observations. The OBIS, like scores derived from other parts of the method (number of MAs, Vigs, etc.) provides a representative number that can be used for quantitative comparisons. Procedures of MAIA can be summarized as follows. Intercoder agreement is sought after each step of analysis:

1. Code enactment videos for MAs, and collate a 'Repertoire of micro-actions'
2. Code the MAs for each enactment into story Vigs, and collate a 'Repertoire of vignettes'
3. While referencing enactment analysis, code interview transcripts
4. While referencing enactment and interview analyses, code drawings
5. Assign OBIS to each enactment from combined understanding of all analyses

CASE APPLICATION OF MAIA

As an instance of use, we describe a study where MAIA was applied. The study investigated the effects of using physical, tangible objects as support for the child's imagination in the context of enactive storytelling. Nine-year old children were asked to enact, using different types of objects (three versions of a toy frying pan, a toy pickaxe and a toy lantern), parts of a story that was shown to them digitally. After each act of the story, they were asked to draw a scene that they enacted. At the end of the study, each child was interviewed for around 10 mins. Figure 1 shows a sample of data collected from the three sources.

The collected video stream was cut up to isolate each enactment of each child. Two coders analyzed each enactment video separately identifying MAs and their timings and recording these in a spreadsheet. Disagreements were then resolved in discussion and a consolidated 'action description' of the enactment was produced. For each enactment, the MAs coded in the consolidated action description were collated to produce the 'repertoire of micro-actions' for that particular enactment. Acronyms such as Flipping (F) or Put in Pan (PiP) were used to represent the MAs.

The two coders then generated the story Vigs by collapsing group of MAs, conferring upon disagreements. The 'repertoire of vignettes' was produced for each enactment. Acronyms were also used for the Vigs, such as Fanning Fire (FNV) and Misflip & Catch (MFCV). A sample of part of the consolidated coding sheet is shown in Figure 2.

After all interviews were transcribed and inserted into the spreadsheet along with the MAs repertoire and the Vigs repertoire, the two coders read and coded the transcripts separately and then together, for the four dimensions of goal, schema, extension and consistency. The two coders also coded the drawings for the three dimensions of character, scene and consistency.

Referencing the videos and the collated spreadsheet, each of the two coders gave an OBIS (by scoring on richness, typicality and consistency) to each enactment on a scale of 0 to 7. Unmatched scores (only 10%) between the two coders were resolved into a score that both agreed was adequately representative. Figure 3 shows the multi-track analysis of two enactments, one with a high and one with a low OBIS. The MA and Vigs descriptions of each enactment provide the coder with a mental model of the enactment while reading the interview and drawing analyses.

We highlight that neither one of the analyses (MAs, Vigs, drawing, interview) should be taken separately. Interpreting them as isolated analyses may at times lead to a very different OBIS than an OBIS from the holistic interpretation. We initially assigned OBIS based on the MA and Vig analyses only and conducted a separate coding session with six external coders who viewed simply the enactment videos. Conducting T-tests on the scores, we found that their given OBIS, as well as our initial ones, had very low correlations with our final OBIS generated based on the integrated view.

The OBIS were entered into SPSS for statistical ANOVA data analysis, together with personality and baseline performance scores as covariates. A significant main effect of object type ($p < .05$) was obtained, allowing us to evaluate how the various versions of the three objects differed in effectiveness in terms of imagination support.

CONCLUSION

We proposed a methodology that makes use of micro-analysis of various data sources to assess imagination as a process of creativity, focusing on embodied storytelling CSEs for children. We described how the approach was used to successfully analyze data from the study of a storytelling environment using physical objects. With further research, we believe MAIA can be adapted for use with other types of embodied creativity support systems for children that are not specific to storytelling.

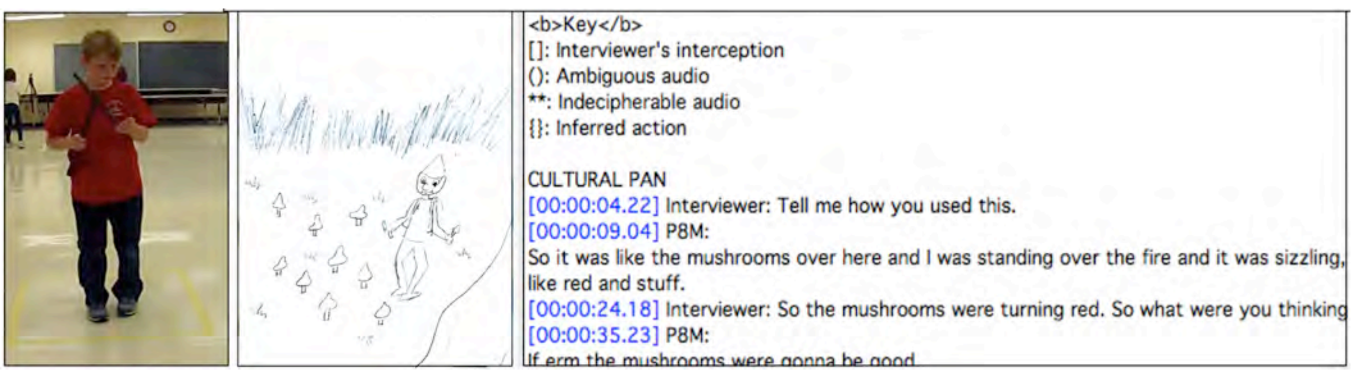

Figure 1. Sample of data sources: Left. Enactment video; Middle. Child's drawing; Right. Interview transcript

High-level Micro-Action Description	Micro-Action Repertoire	Vignettes
Child is in kneeling position throughout but with a more relaxed feeling than in Act 1. At the beginning of the enactment, he glances over at the female subject across the room using the cultural object. She was 'adding condiments' to her pan (PiP). He seems to follow her action. Throughout the enactment, he adds condiments, shakes the pan (small motion side to side) (HSS) and repeats the cycle. He generally looks at the pan and the cooking activity he is doing. At about 1:30 he seems to get bored and disengages. At 1.56, he seems to do a single flip (F) after glancing at the girl again (she was doing large flips).	1. Put in Pan (PiP); 2. Horizontal Shaking Side to side (HSS) 3. Flipping (F)	**Summary: BCV** (Basic Cooking Vig), **AIV** (Add Ingredients Vig) Child does BCV constantly with a lot of AIV. He does mostly SoF, adding a few subtle HSS at times and a single F toward the end. He adds ingredients throughout from both sides.

Figure 2. Sample of MA/Vig analysis spreadsheet

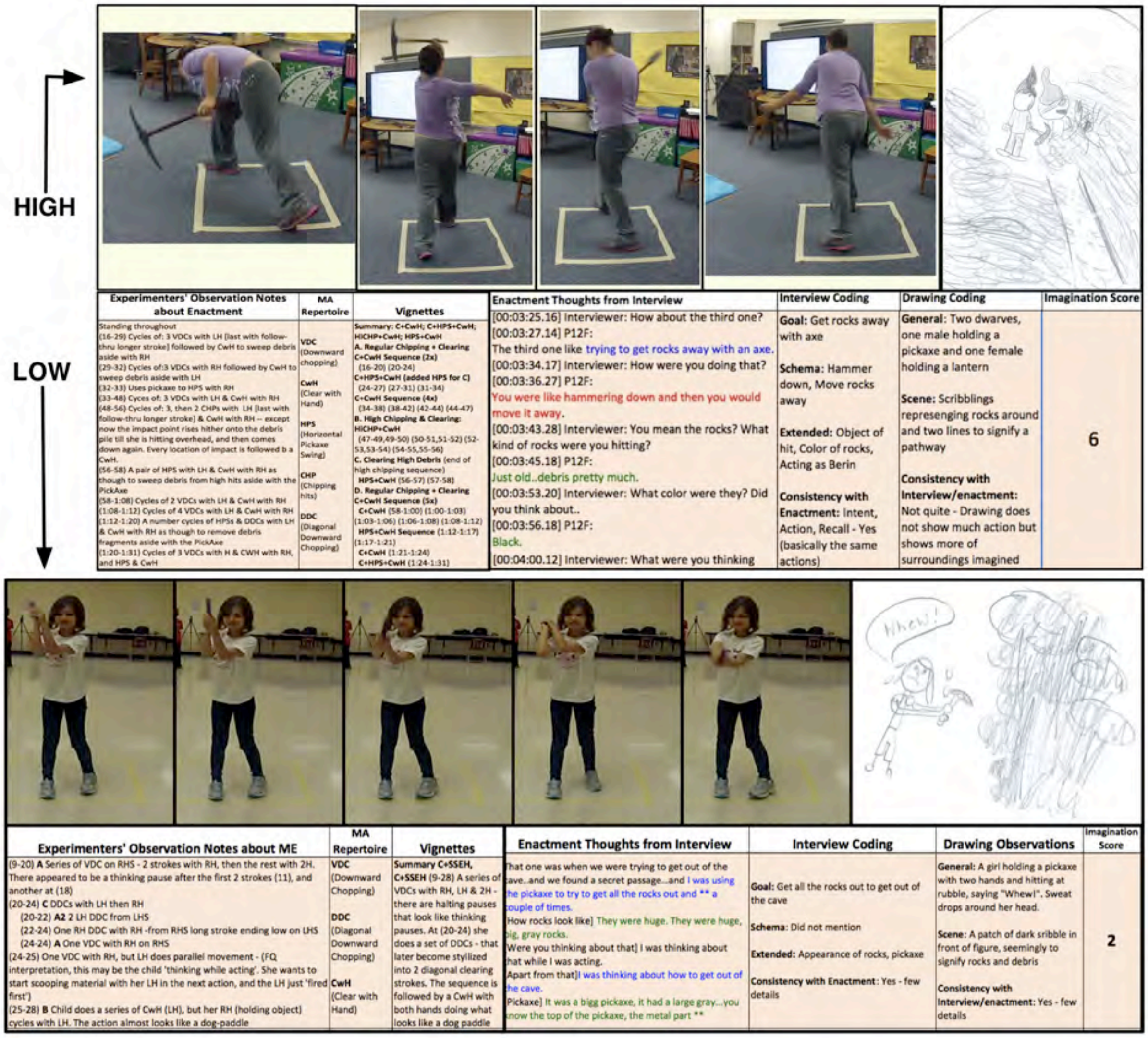

Figure 3. Sample analyses of enactments with a High and a Low OBIS. Top left part of each block shows frames from the enactment videos taken at intervals. Top right parts show the child's drawing of the enactment. Bottom parts show the MA, Vig, interview and drawing analyses, and the intercoder-assigned OBIS.

REFERENCES

1. Ryokai, K. and J. Cassell. *StoryMat: A Play Space for Collaborative Storytelling*. in *CHI*. 1999. ACM.
2. Polazzi, L. *The Pogo active tools: Narrative logic, time and space in children storytelling. Interaction without frontiers: Joint AFIHM-BCS Conf. on HCI,* 2001.
3. Fontijn, W. and P. Mendels. *StoryToy the interactive storytelling toy*. in *2nd Intl Wksp on Gaming App in Pervasive Comp. Environments at Pervasive*. 2005.
4. Sherman, L., et al. *StoryKit: tools for children to build room-sized interactive experiences*. in *CHI'01 extended abstracts* 2001. ACM.
5. Vygotsky, L.S., *Mind in Society: The Development of Higher Psychological Processes*1978: Harvard U.
6. Vygotsky, L.S., *Imagination & Creativity in Childhood.* J. Russian & E. European Psych., 2004. **42**(1): p. 7-97.
7. Pelaprat, E. and M. Cole, *Minding the Gap": Imagination, Creativity, and Human Cognition.* Integrative Psychological and Behavioral Science, Springer, 2011. **45**: p. 397-418.
8. Finke, R.A., T.B. Ward, and S.M. Smith, *Creative Cognition*1992, Cambridge: MIT Press.
10. Boden, M., *The Creative Mind: Myths and Mechanisms*1990, London: Basic Books.
11. Fauconnier, G. and M. Turner, *The Way We Think. Conceptual Blending and the Mind's Hidden Complexities* 2002, New York: Basic Books.
12. Markey, F.V., *Imagination.* Psychological Bulletin, 1935. **32**(3): p. 212.
13. Loke, L., et al., *Understanding movement for interaction design: frameworks and approaches.* Personal & Ubiq. Comp., 2007. **11**(8): p. 691-701.
14. Andrienko, G., et al., *A conceptual framework and taxonomy of techniques for analyzing movement.* J. of Visual Languages & Comp., 2011. **22**(3): p. 213-232.
15. Nemirovsky, R., M.L. Kelton, and B. Rhodehamel, *Gesture and Imagination: On the Constitution and Uses of Phantasms*, 2011, http://mathmoves.org/sites/default/files/gestures_and_imagination_plates_and_figures_embedded.pdf.

Evaluating Creativity Support in Co-Design Workshops

Graham Dove
Centre for HCI Design
City University London
graham.dove.1@city.ac.uk

Sara Jones
Centre for Creativity in Professional Practice
City University London
s.v.jones@city.ac.uk

ABSTRACT
Participatory, co-design and creativity workshops can lead to more useful, usable and innovative systems design. However, evaluating the effectiveness of the creativity support provided by different technologies and workshop techniques is challenging. This is especially so when evaluation takes place during the workshop and maintaining a creative atmosphere is important. In this paper we briefly outline the development of one simple method of evaluation we have designed whilst studying the use of information visualizations within generative design workshops. Here we discuss how reflective postcards are used to replace questionnaires as a way to collect participants' responses.

Author Keywords
Evaluation, Creativity Support, Participatory Design

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Human Factors; Design;

INTRODUCTION
Participatory approaches to human-centred design, characterized by the active involvement of users and other stakeholders, can lead to more useful and usable systems [7]. Through practices such as co-creation [9] and creativity workshops [6], it has also been shown that such methods can be an effective way to discover novel requirements for complex socio-technical systems and design future experiences for their users. A key aspect of these approaches is the requirement for designers to provide the tools and facilitation skills that elicit participants' possibly latent creativity. It is therefore crucial that as far as possible any such workshop retains an atmosphere that is relaxed, supportive, engaging and playful.

When undertaking academic research to study the effectiveness of particular technologies, techniques or activities, it is sometimes important to collect evaluation data from participants during the workshop itself. This creates something of a conflict as stopping generative activities to ask participants to complete questionnaires serves only to highlight academic concerns. This can also draw attention to any possible concerns participants have that they themselves are being judged, which can be a cause of anxiety. Because of this we have sought evaluation methods that become part of the workshop's creative activities. In this paper we will briefly outline the development of a simple method that, whilst not entirely novel in its intentions, is one we hope will be of interest. We have found it useful in capturing evaluation data similar to that in questionnaires but using a form factor that is more appropriate to the workshop context. Here, individual postcards containing prompts designed to capture participants' reflections are used to assess selected aspects of the workshop up to that point. We do this in order to evaluate the support participants feel a particular technology or technique has provided them for their role in the workshop activities they have just undertaken. In the following sections we will first provide some background to our wider research, before discussing the stages that led to the development of this evaluation method. We will then describe how it has been used in practice and close with a brief discussion of its effectiveness.

BACKGROUND
In our research [4], we are investigating ways information visualization can be used in conjunction with generative tools and creativity techniques to support participatory design research. This is in response to the large amounts of data organizations now hold, following the movement of services and transactions online and as a result of the increasing ubiquity of computing systems. These data can be reused, offering an opportunity to create innovative products and services, but the contexts surrounding these data present all the difficulties associated with 'wicked problems'. To address these challenges we seek to provide participants with a combination of tools, techniques and support that enables them to better understand data, explore current context and imagine possible futures. By information visualization we are referring to the graphical representation of data. This will often, although not exclusively, be interactive. By generative tools we are referring to methods whereby stakeholder representatives are provided with the materials and techniques, such as those needed for making collages, to help them generate new ideas that reveal requirements or inspirations for design.

In our research we need to evaluate the introduction of information visualizations into different types of generative workshop activity. In addition we need to evaluate how effective different generative activities are in helping participants and designers gain a shared understanding of data. We must also evaluate the effectiveness of different styles of information visualization and their appropriateness within our workshop activities. Each of these evaluations takes place within the context of participatory design workshops. Following Warr and O'Neill's description of design as a social creative process [10] we frame these evaluations in three parts, addressing the creative process, the creative product and the creative person. This is described further elsewhere [4]. In the remainder of this paper we will outline our method for assessing participants' self-reported evaluation of how effectively they feel a particular technique or technology has supported them in their role as a creative person within an ongoing workshop.

DEVELOPING THE REFLECTIVE POSTCARDS

Stage 1: Separate Questionnaires

In an earlier piece of research, evaluating the support a large-scale interactive visualization of student satisfaction data provided for collaborative ideation [3] we had used three separate questionnaires to address system usability [1], creativity support [2] and insight support. The third of these was a questionnaire we developed ourselves based on previous work outlining the nature of insight as discussed in the visualization community [8] and analysis of how such insights are acquired whilst visually exploring data [11]. Whilst each of these questionnaires was successful in addressing the concerns it covered, the process of completing them all was a chore for participants. This had a negative impact on the quality of the responses to subsequent open questions we asked to probe participants' qualitative experiences.

Stage 2: A Single Questionnaire

As a result of this, when we undertook a study comparing the effectiveness with which two different styles of information visualization provided stimuli for ideation, we decided first to separate out the usability evaluation and then to combine the creativity support and insight support questionnaires into one. Here we wanted to design a short and simple questionnaire that would quickly address participants' responses to the most salient aspects under investigation. This would then form a small but nonetheless significant aspect of our overall evaluation plans. The result was a seven-part questionnaire that used a Lickert scale rating, ranging from 1 strongly agree to 5 strongly disagree. The first four statements in it are derived from the Creativity Support Index [2] and the final three from the insight support questionnaire we had developed based on visualization literature [8,11]. The questionnaire statements are listed below:

1. I was very engaged and absorbed using the visualisation. I enjoyed it and would do it again.
2. I was prompted to generate ideas that were new and varied.
3. I was able to work together with others easily.
4. I felt able to explore many different options, ideas or outcomes.
5. I could easily identify relationships and patterns in the data that contributed to new ideas.
6. It was easy for me to gain an overview of the data using the visualization.
7. I was able to combine my existing knowledge with insights from exploring the visualization to generate ideas that I had not previously considered.

This questionnaire was successful in the context of a design experiment as it captured responses to our main concerns. However, here it was being presented at times when there was a clear and intentional break in the flow of creative activities, and where a change of atmosphere was both appropriate and desired.

Stage 3: Reflective Postcards

This would not be the case in a workshop in which end users and other stakeholders were participants. Here we would want to keep the focus of those participants away from our academic concerns. We would not want them to feel they were being assessed and we would not want to break the flow of generative creativity. Here we would need an alternative format. The criteria we had for an evaluation method were as follows. First, it should feel personal, encourage reflection and allow for creative responses. Second, it should be relatively short but directed at answering particular areas of concern. Third, it should fit into the activities of the workshop without changing the atmosphere or drawing participants' attention to assessment. Fourth, it should use a mechanism that would be familiar to all participants. Finally it should be able to capture data replacing the Lickert scale responses and also the open questions we had asked in previous questionnaires.

Gaver & Dunne's use of cultural probes [5] tells us about the effectiveness of well-designed prompts and intriguing artifacts in eliciting responses from people, and this was a source of inspiration in our decision to use postcards as a medium to collect evaluation data. Postcards are individual artifacts that limit the space in which responses can be written but which are flexible enough to provide the opportunity for creativity. They have a form factor that is both familiar to people and evocative of sharing. They also suggested a playful means of collection to complete the activity, and so we made a small red postbox for participants to 'send' us their reflections.

USING THE REFLECTIVE POSTCARDS

The workshop for which the postcards idea was developed was held as part of a research project undertaken in conjunction with E.ON [4], a major energy provider, in which we are investigating possible new products or services that could be developed using the data generated by smart home technologies. Within this project a sophisticated model of typical energy consumption patterns has been developed. We used the data generated by this model to build interactive information visualizations that provided stimuli during a pair of workshop activities. In the first, participants worked in small groups to create collages describing different aspects of the household they imagined might be represented by the energy consumption data. In the second, these data were further explored to complete a competition entry outlining ways in which the imagined household could be smarter in their energy use.

These imaginary households and the contexts of their energy consumption behaviour would later be used as inspiration in service design activities, but we wanted to evaluate participants' responses to using the information visualizations immediately following the activities in which they were being used. In order to achieve this each participant was given three separate postcards. Each postcard had a different reflection prompt written on it that we asked participants to respond to. We chose to use prompts that asked for reflection rather than standard open questions because we felt that this approach would encourage participants to think critically and discuss both what had been effective and also what hadn't worked. These prompts were derived from the questions we had used in the earlier design experiment.

Reflection Prompt 1

The first prompt addresses engagement and collaboration, similarly to statements 1 and 3 in our earlier questionnaire.

"Please reflect on your involvement in the previous two activities. Write a few sentences thinking in particular about how engaged you were, how absorbed or distracted, and how easily you feel you worked with other members of your team. Try to think about the extent to which the technology helped or hindered you in this regard."

Reflection Prompt 2

The second prompt addresses idea generation, exploration of alternatives, and the ease with which participants could utilize their knowledge and experience. This is similar to statements 2, 4 and 7 in the questionnaire.

"*Please reflect on how you used the data visualization to first create your household and then to devise competition answers. Write a few sentences, thinking in particular about how easily you were able to explore possible options and come up with different ideas. Did you use your prior knowledge as well as the information shown? And how easy you found it to relate that prior knowledge to the data?"*

Reflection Prompt 3

The third prompt addresses participants' ability to gain an overview and to identify relationships and patterns within data. This is similar to statements 5 and 6 in the questionnaire.

"Please reflect on your understanding of the information contained in the data visualization. Write a few sentences, thinking in particular about how easily you managed to gain an overview of what was represented. Also think about how quickly you grasped what the information meant, did you spot clear patterns and relationships or did you find it confusing? Did it prompt you to think of ideas you had not previously considered?"

EXAMPLE REPONSES TO REFLECTIVE POSTCARDS

Participants responded well to the postcards, taking the time and effort to provide considered responses addressing both positive and negative factors. We have listed example responses to each of the postcards below.

Reflective Postcard #1

"It was easy to work with the group, we were open to each others opinions. Technology was useful for us to investigate our views and to help discussion. The display of the information was interpreted differently by others but this helped with discussion."

"I felt engaged and absorbed with the tasks and comfortable working with the other members. Some of the information in task 1 was a little overwhelming. The technology was very useful."

Reflective Postcard #2

"The iPad data visualisation was very useful as it made it surprisingly easy to look at each piece of data and also caused the data to be better laid out. I could also use it with my own knowledge which I had to do for the first task."

"Did use prior knowledge and so did other team members. Needed to focus back on house and empathise what they were like. iPad data didn't really contribute to ideas."

Reflective Postcard #3

"Definitely. You had a broad overview and you could drill down to get clearer answers. This interactivity flowed very well and really demonstrated well how this family behaved."

"It was easy to get an overview about each group of data due to how it was laid out and that made it very easy to compare the data and come to assumptions about it."

ANALYSIS OF REFLECTIVE POSTCARDS

Our analysis indicates that Reflective Postcard #1 and Reflective Postcard #3 successfully replaced the equivalent questionnaire items and elicited responses relevant to our concerns. In the case of Reflective Postcard #1 all 13 participants responded to the engagement aspect and 12 to collaboration. Reflective Postcard #3 gave us 12 responses to identifying patterns and relationships and 9 regarding overview. However, for Reflective Postcard #2, whilst 7 of the 13 participants responded to the prompt regarding use of their existing knowledge, only 2 addressed exploring alternatives and just 1 idea generation. We characterised each of these responses as being either positive or negative. The results are shown in Table 1.

Evaluation Factor	Reflections	
Engagement (Q1, P1)	+13	-0
Collaboration (Q3, P1)	+12	-0
Generating Ideas (Q2, P2)	+0	-1
Exploring Options (Q4, P2)	+1	-1
Building on Existing Knowledge (Q7, P2)	+7	-0
Patterns & Relationships (Q5, P3)	+10	-2
Overview (Q6, P3)	+8	-1

Table 1: Analysis of Responses on Reflective Postcards

The Reflective Postcards are not designed to capture data in the depth required for a systematic qualitative analysis. However they do provide responses similar to those from open-ended questionnaire questions. This helped with Reflective Postcard #2, where responses had not referred directly to the subjects posed in the prompt. Here, participants took as much consideration as they did with the other postcards, but we found they were taking the opportunity to provide us with suggestions for improvement or more generally helpful feedback. For example one participant wrote:

"Very helpful. Couldn't do it without. Some minor improvements (red for bad?). Took knowledge to use it. May be difficult for non-expert."

Whilst another participant used it as an opportunity to relate the workshop to the wider trial of smart home technology that our participants are a part of:

"To an extent, not having full Greenwave socket data on major appliances. Much of it was a 'guesstimate'."

DISCUSSION

The responses to Reflective Postcards #1 and #3 suggest that they can be an effective replacement for questionnaires within a workshop. However the evidence from Reflective Postcard #2 suggests further refinement is needed to explore their limitations. It could be that participants did not specifically address the question of idea generation in Reflective Postcard #2 because the activities undertaken with the visualization were not obviously ones requiring divergent thinking or rapid idea generation and instead participants took the opportunity to share more general thoughts. Or it may be that this prompt, in addressing three separate concerns jointly, was too ambitious or simply not clear enough. Further study and improved piloting will help identify and militate against similar problems in future.

These postcards offer us more than a simple replacement for questionnaires though. Their format is flexible and affords participants the opportunity to be creative in their responses. They are also informal and do not draw attention to notions of assessment. In their use we are investigating participants' reflections regarding the support particular techniques or technologies can provide them in their role as a creative person undertaking workshop activities. We need to do this in a way that maintains the atmosphere of the workshop and the postcards seem to meet the requirements to achieve this. So far, participants have responded to the postcards with a good level of enthusiasm and they seem to like their somewhat homemade styling. In order to encourage this aspect we are continuing to adapt the postcards and develop them as artifacts by adding images and providing envelopes addressed to us personally, in case participants wish to take them away, reflect further and return them to us by post. As a method of evaluation, the postcards are simple and yet effective in assessing the support provided to participants as the creative person.

ACKNOWLEDGEMENTS

We would like to thank Alison Duffy for her help.

REFERENCES

1. Brooke, J., SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189 (1996) 194.
2. Carroll, E.A, Latulipe, C., Fung, R., Terry, M., Creativity Factor Evaluation: Towards a Standardized Survey Metric for Creativity Support. In *Proc. C&C '09,* ACM Press (2009), 127-136
3. Dove, G. CityVis A Visualization Tool to Support Exploratory Creative Ideation on City University Data. MSc Dissertation (2011) http://www.grahamdove.com/papers/cityvis.pdf
4. Dove, G., Jones, S., Dykes, J., Duffy, A., and Brown, A., Using Data Visualization in Creativity Workshops. *Submitted to C&C 13* (2013) http://www.grahamdove.com/papers/candc2013.pdf
5. Gaver, B., Dunne, T., & Pacenti, E. Design: cultural probes. *interactions*, *6*(1), (1999), 21-29. ACM
6. Jones, S., Maiden N.A.M., Karlsen K., Creativity in the Specification of Large-Scale Socio-Technical Systems in *Proc. CREATE 2007* (2007) 41 – 46
7. Muller, M.,J., Kuhn, S., Participatory Design. In *Communications of the ACM*, 36, 6 (1993), 24-28
8. North, C., Towards Measuring Visualization Insight, *Computer Graphics and Applications, IEEE* 26,3 (2006) 6-9
9. Sanders, E.B.N, Information, Inspiration and Co-creation. In *Proc. 6th International Conference of the European Academy of Design*. (2005)
10. Warr, A., & O'Neill, E. Understanding Design as a Social Creative Process. In *Proc. C&C '05* ACM Press (2005) 118-127.
11. Yi, J. S., Kang, Y. A., Stasko, J. T., & Jacko, J. A. Understanding and Characterizing Insights: How do People Gain Insights Using Information Visualization? In *Proc. BELIV '08*. ACM. (2008) 4-10

Supporting Creativity in Performance Lectures

John VH Bonner
School of Computing and Engineering
University of Huddersfield
HD1 3DH
j.v.bonner@hud.ac.uk

David Peebles
School of Human and Health Sciences
University of Huddersfield
HD1 3DH
d.peebles@hud.ac.uk

ABSTRACT

This paper presents a reappraisal of how conventional lectures could meet the expectations of 'digital native' students. Our approach is to consider how digital media could be integrated further into the lecture by transposing knowledge through artistic as well as pedagogical means to form a 'performance lecture'. We began by developing a design and evaluation framework to foster new creative thinking and also to assist as a structuring tool. The framework draws on the disciplines of interaction design, cognitive psychology and performance art. The framework consists of five elements, four design production parameters and a performer/audience interaction model. The production parameters consist of modal expressions, spatial design, temporal flow, and audience engagement and spectatorship. Connecting the five elements together are rhetorical considerations. This paper presents the outline structure of this framework and concludes with some findings from an evaluation of framework effectiveness to help in creatively developing performance lectures.

Author Keywords

Performance lectures; interaction design; cognitive psychology; performance arts; audience engagement

ACM Classification Keywords

H.5.5 Information interfaces and presentation: Sound and music computing
H.5.m. Information interfaces and presentation: Miscellaneous.

General Terms

Human Factors; Design; Measurement.

INTRODUCTION

The motivation for this research is to challenge and reinvigorate the conventional lecture and discover what a lecture could become by thinking creatively about the deployment of digital media technology, not as a supplement to the lecture but as an intrinsic part of the learning experience. Learning should not be based on what technology can do but on how audiences can learn within a digital media environment [18].

A 'lecture-as-performance' view helped foster alternative delivery mechanisms such as combining and mixing conventional oration with story-telling, spatially distributed live and recorded audio and music, text and graphical animations, multiple video projections onto different walls, even the floor and ceiling. While many of these suggestions might provide an audio-visual spectacle, our interests lie in exploring if this approach could also be pedagogically successful?

A performance lecture is therefore a new concept and this paper describes the early embryonic stages of its development, in particular how to manage what could be an unwieldy creative process. As examples of performance lectures and this particular hybrid approach do not yet exist, we decided first to produce a development framework that would act as an instrument for evolving lectures into performance lectures. Our methodological approach was inspired by related research and artistic work using mixed reality described by Benford and Giannachi [2] who refer to the use of 'sensitizing concepts' as a means of providing guidance and analytical direction for exploratory fields of creative and evaluation study.

PROPOSED DEVELOPMENT FRAMEWORK

We view a performance lecture as a hybrid artefact stemming from three disciplines: cognitive psychology, performance art and interaction design. As one of the goals is to ensure performance lectures engage audiences through novel methods of dissemination, it's important that audience interaction is considered at the cognitive level. For example, there are known cognitive processing strategies of text and visual to ensure effective learning if delivered concurrently [7].

Interaction design offers a range of analytical methods and has been used to model information flow between the performer and audience. Performance art provides a historical context, creative inspiration and critical argument on the nature of performance. By bringing these disciplines

CHI'13, April 27 – May 2, 2013, Paris, France.

together, our hope is we might successfully combine artistic and pedagogical perspectives.

The framework consists of six elements. These include four production parameters: modal expressions, temporal flow, spatial design and audience engagement which are collectively managed through the rhetorical aspects of the performance lecture. The intended thoughts, ideas and meaning of the performance lecture are constructed by independently and collectively considering these five elements. Through this process a performance lecture is produced and articulated as a notated interaction model between the performer and audience. Each performance lecture has its own interaction model and because of this the effectiveness or experiences can be assessed against the audience's recall of the lecture.

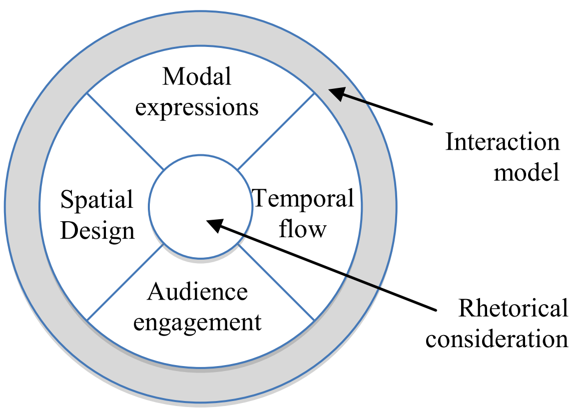

Figure 1. Framework for design and evaluation of performance lectures

Because conventional rules and rituals of a lecture are being re-appraised, rhetorical considerations are required to manage expressive and creative latitude. The art of good audience engagement depends on the performer being able to make appropriate content and delivery decisions about what to make explicit, inferential, peripheral or redundant. Other decisions come into the frame such as how far to depart from a 'mono-logical and mono-vocal narrative' [1]. However, as we intend to use digital media in more creative ways, this 'performer' role may not be so distinct. Lecture audience expectations will need to be carefully managed as perceived integrity often lies in social conformity and audiences may become skeptical if the event appears to be intellectually superficial.

It is this blurring of the rhetorical stance between a lecture and a performance lecture that is intriguing as a production problem. Without rhetorical consideration, an audience may find it difficult to understand their role as a listener or spectator. Thus an important aspect of this research will be to develop mechanisms to assist the audience to 'decode' the performance lecture; to encourage active analytical listening and improve the aesthetic experience when conventional narrative structures may not exist.

The four 'quadrant' elements in Figure 1 provide opportunity for divergent thinking about content and delivery. The common modal expression for a lecture is oration. Our intention with this framework was to allow other modalities to be critically considered as part of the performance lecture environment. Music can evoke physiological as well as strong emotional feelings in the listener. Music fosters strong but imprecise emotions and our research suggests that its role will be limited to mood setting. However, soundscapes would appear to be more flexible. Soundscape is a term defined by Schafer [9] to refer to an auditory landscape which can consist of natural and/or artificially created sounds. Soundscapes have very powerful 'referent' qualities and can be used to alter or affect human behaviour.

Moving images such as animation or video have also been used to augment and supplement key points in lectures. The ubiquity and acceptance of the projected images, we think, provides an opportunity to use images differently. Consideration can be given to foreground and background roles, or providing new ways of demanding attention or perhaps purposefully deflecting from it.

The spatial design in which the performance takes place also offers creative latitude. Stage placement and delineation also offer creative possibilities for re-negotiating physical cues between the audience and performer. Multi-screen displays are used in the entertainment industry and can be used as a form of stage. Although large visual displays are used at public events, there has been little research on how the audience experience of multi-screen productions can be creatively exploited beyond visual wallpaper although some research have been carried out in an educational context [3] Location-based audio systems have been used to enhance learning at tourist attractions [5] or for artistic purposes. However, we are not aware of any research work where spatially distributed audio has been used for educative purposes.

Deeper audience engagement will be achieved by encouraging active or deep listening through cognitive engagement. One way of doing this is by adhering to the listener's pattern of expectations and also playing with the cognitive functions of listening such as 'differencing' which selects or de-selects what we listen to and pattern recognition which allows us to identify and tune into certain sounds.

Finally, the outer layer of the framework, the interaction model, allows a performance lecture to be structured and annotated not just in terms of content but also to make explicit representations of action and thought. The model borrows elements of activity theory [4, 8] and Laurel's Elements of Qualitative Structure [6]. Activity theory was important to consider because it brings together activity using a wider social and organisational context combining this with 'consciousness'. This cognitive engagement needs to be explicitly articulated and this is achieved through ‘objects’. The term 'object' is an important concept in AT and differs in meaning to its typical understanding in HCI or design. It was important that this model encompassed abstract and concrete representations of meaning and understanding in both the performer and audience. Thus, effectiveness of the performance lecture can then be measured by identifying parity between commonly shared artistic and pedagogical objects from the performer to the audience.

To help contextualize the abstract description of the framework and understand how it can be used to creatively design and evaluate future performance lectures, the framework was used to design a three-screen (triptych) video and the interaction model and then used to evaluate the effectiveness of the video presentation as an innovative medium. The triptych video spatially and temporally spreads video and audio content across three large displays; thus forcing the viewer to actively ‘assemble’ a narrative and meaning from the fragmented elements. The video consisted of a reportage description of a manufacturing organisation.

It must be stressed the triptych video was not a performance lecture but could be a potential component of future performance lectures. The aim at this stage was simply to validate the framework and evaluation model.

An evaluation study was carried out with three objectives. 1) To explore how much of the pedagogical and artistic intent embedded into the triptych video was successfully conveyed and what type of audience experiences were gained. 2) To assess the framework as an instrument for developing performance lectures. 3) To assess the impact a post-hoc evaluation approach has on future creative thinking for the design of performance lectures.

12 students volunteered to watch the 12-minute video. All subjects viewed the video alone except for the experimenter. Eye gaze during the video along with subsequent interview were video recorded for data analysis purposes. Subjects were undergraduate students predominately between the age of 21-30 studying either psychology or computing.

After watching the video, subjects were interviewed using structured and semi-structured questions to elicit recall accuracy of key events and scenes, personal experiences of any cognitive load, how they felt their attention had been guided using the multi-screen design and finally an overall appraisal of their experience in terms of engagement, entertainment, and educative value. The results from the questionnaire and interview transcripts were then reviewed and compared against the intended performance objectives of the triptych video.

The feedback gained from this study revealed subjects failed to comprehensively interpret the video. Our explanation for this related to the contextual setting of the study. Subjects were not given any form of preparation or introduction to the video and the subject matter was unfamiliar to them. The lack of an introduction was quite deliberate because we were anxious not to bias their expectations. We were anticipating some disinclination towards the video; however, the limited lack of awareness to the mixed modal themes was still surprising. In contrast, the video was also shown to some of the management team from the organisation the video was about. Their responses were diametrically opposite to the experiment subjects. They were amused by the unusual juxtaposition and fragmentation of scenes and knew exactly where each shot had been taken and thought the video accurately conveyed the company but not in ways they would have anticipated.

Findings from the study have helped inject caution into the way novel approaches to audience engagement are introduced and has highlighted the need for a stronger audience preparation for their spectator experience. With this in mind, we are beginning a new comparative study focusing principally on different methods of audience priming.

Conclusion

By iteratively moving through this framework by radiating from the centre to the perimeter and then back to the centre, it should be possible to derive innovative performance lectures that offer engaging audience experiences using novel approaches to content delivery.

The process should begin with rhetorical considerations. However, we feel defining these ‘considerations’ still requires deeper clarification. Our intention is to continue with this iterative process by developing small parts of a performance lecture until the component elements of the framework stabilize. Once this occurs, performance lectures can then be confidently designed with purposeful artistic and pedagogical intent imbedded into them.

REFERENCES

1. Bagley, C. (2008). "Educational ethnography as performance art: towards a sensuous feeling and knowing." Qualitative Research 8(1): 53-72.
2. Benford, S. and G. Giannachi, Performing Mixed Reality. 2011, Cambridge, Massachusetts: MIT Press.
3. Bligh, B. and K. Lorenz, The rhetoric of multi-display learning spaces: exploratory experiences in visual art disciplines. International Journal of Media, Technology and Lifelong Learning, 2010. 6(1): p. 7-27.
4. Bødker, S., Through the interface: a human approach to user interface design. 1991, New Jersey: Lawrence Erlbaum.

5. FitzGerald, E., et al., Guidelines for the design of location-based audio for mobile learning. International Journal of Mobile and Blended Learning, 2011. 3(4): p. 70-85.
6. Laurel, B., Computers as theatre. 1993, Reading, Massachusetts: Addison-Wesley.
7. Mayer, R.E., Multimedia learning: Are we asking the right questions? Educational Psychologist, 1997. 32: p. 1-19
8. Nardi, B.A., Some reflections on the Application of Activity Theory, in Context and Consciousness, B.A. Nardi, Editor. 1996, MIT Press: Cambridge, Massachusetts.
9. Schafer, R.M., The Soundscape: Our Sonic Environment and the Tuning of the World. 1977, Rochester, Vermont: Destiny

Searching to Measure the Novelty of Collected Ideas

Rhema Linder
Interface Ecology Lab
Department of Computer Science and Engineering Texas A&M University
rhema@ecologylab.net

Andrew M. Webb
Interface Ecology Lab
Department of Computer Science and Engineering Texas A&M University
andrew@ecologylab.net

Andruid Kerne
Interface Ecology Lab
Department of Computer Science and Engineering Texas A&M University
andruid@ecologylab.net

ABSTRACT
Visual representations of ideas are valuable for creative thinking and expression. Prior research on design and information-based ideation has assessed novelty in creative products as the inverse of the frequency that an idea or visual element occurs in the complete space of responses. In controlled experiments, frequency has previously been calculated in reference to the set of ideas collected by all participants (corpus). Experimental conditions restricting the space of possible elements resulted in overlap between participant responses, yielding a range of frequencies. Alas, in field investigations the space of possible elements is unrestricted, resulting in little overlap of ideas, and thus mostly a single frequency ($1/N$) of collected elements.

We introduce a new method that uses web search to measure the novelty of individual ideas. Instead of using the local corpus directly to calculate frequency, we use the number of results from web searches generated from elements in the corpus. Our implementation uses Googles reverse image lookup to determine the popularity of images. We compare results with those derived via prior methods.

Author Keywords
novelty, information-based ideation, evaluation metrics

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION
Visual representations of ideas are valuable for creative thinking and expression [1, 4]. Prior research on design and information-based ideation has assessed novelty in creative products [8, 5, 9]. In researching creativity support environments, researchers often label user authored media by hand to extract data from experiments and investigations in the field. As the amount of user authored media increases, researchers have more impetus to invent automated techniques for measuring components of creativity.

Information-based ideation (IBI) is the process of having new ideas while working with information [5, 11]. In *information-based ideation tasks*, a person searches, collects, organizes, and thinks about information to answer open-ended questions, such as planning a vacation, deciding on a thesis topic, or designing an innovation. In IBI tasks, people author collections of information and media elements from sources including the Web to represent ideas relevant to the task at hand. In this methodology, each collected element is an individual "answer" or idea developed in response to an IBI task, which one can use to measure components of creativity.

Novelty, one component of creativity, is the uniqueness of an idea. Researchers have computed novelty as statistical infrequency in laboratory experiments [8, 9, 11]. We refer to the prior method for calculating elemental novelty as *corpus-based novelty*. In corpus-based novelty, a collected element has high novelty if it appears rarely over the set of all collections in an experiment (the corpus). However, this metric only produces a rich range of values when the number of possible elements an author can collect is constrained by experimental conditions. When users collect media from the internet using self authored queries, the number of possible elements to choose can include any online content. An element collected in such unconstrained conditions is likely unique among participants, but may be commonly or uncommonly found online.

Search-based novelty uses the number of search results from a query generated from a media element to measure its novelty, producing a rich range of values. Computing search-based novelty is an algorithmic process, which is faster than manual solutions that require human raters. Once a set of sample of media is used to create a novelty function, it not require on a corpus of experimentally collected data to calculate novelty.

We begin with a discussion of prior work. Next, we introduce search-based novelty. We show empirical evidence to validate this metric in the context of images and Google Image search. We conclude by discussing implications for design and future work.

PRIOR WORK
Prior work has addressed measuring creativity, ideation process, and the experiential ratings of creativity for evaluating creativity support environments.

Carroll et al. produced a range of questions for evaluating creativity support tools [2]. Self reported data provides insight about participant experiences. Our approach measures creative products, which contain attributes that we can measure, rather than a creative processes or experience.

Dow et al. used a variety of metrics to evaluate the efficacy and creativity of ad prototypes [3]. Instead of coding attributes of the creative products as a measure for novelty, they used Mechanical Turk to measure similarity for each ad in the corpus of created ads. To verify, Dow et al. also employed a panel of experts and used the click through rate of ads to measure ad value. Instead of measuring similarity as pairwise, both corpus-based and search-based metrics described in this paper use statistical infrequency to describe novelty.

Shah et al. review and posit measures and processes for evaluating the effectiveness of methods for generating ideas in the context of engineering design: the intersection of utility and novelty [8]. To assess novelty in the context of a highly controlled experiment, they prescribe aggregating ideas present in experiments into categories. The more often an idea occurs across participants, the less novel it is. Calculating novelty with this approach is easy, but requires many hours of work. Similarly, we use uncommonness to measure novelty, but our analysis is automated.

Webb and Kerne et al. designed a laboratory experiment evaluating components of creativity of participant authored information compositions [5, 11]. *Information composition* is a medium that affords collecting and organizing ideas as text and image bookmarks, helping people perform IBI tasks. Participants collected images from a preselected set of search queries. Under these conditions, nearly all of the images collected were collected by multiple participants, creating significant overlap. We report a field study where each participant collected images with little overlap.

Webb and Kerne use Equation 1 to calculate the image novelty of an individual element image e in a set of all image collections C is 1 divided by the number of sets in C where e is present:

$$enov_i(e, C) = \frac{1}{|c \in C : e \in c|} \quad (1)$$

As elements in a corpus become more common, the value of $enov_i(e, C)$ becomes smaller, indicating that the element is less novel. Also note that the novelty of an element e is dependent on the corpus of participant respondent collections C. Calculating the inverse popularity of elements among a corpus closely follows principles that make IDF useful in information retrieval.

Inverse Document Frequency (Equation 2) highlights rare terms in documents by giving high weights to terms that occur in fewer documents over a corpus [7]. The IDF of a term t depends on the number of times that term appears in documents in D.

$$idf(t, D) = \log \frac{|D|}{|d \in D : t \in d|} \quad (2)$$

Figure 1. Example images from least to most novel. Images $inov$ scores ranging from 0 to 1 and e_n from $135,000$ to 1. The Facebook logo is the least novel. Next, an image of Apple TV2, then a stock photo, a photo of a self checkout system, a us quarter back, and a diagram of power grid.

SEARCH-BASED NOVELTY

Search-based novelty measures the uniqueness of a media element with the number of pages it appears on the Web. We show an example implementation using images as media elements in the next section. The process for implementing a search-based novelty metric includes four steps:

1. Collect a set of sample of media elements.

2. Generate search queries for each element algorithmically.

3. Perform searches with each generated query, collecting the number of search results for each element.

4. Create a function that maps novelty inversely to the number of search results produced by an element.

Search-Based Image Novelty

We present our implementation, *search-based image novelty*, which uses Google Image searches to measure the novelty of images found on the Web.

Google Images indexes almost all images on the Web, allowing one to search for images using text or image queries. Google Images combines very similar images, making the search invariant to resolution and small visual perturbations.

First, collect a set of sample images C (1). We used a set of 3,579 images from the information compositions from the field study described in the next section. For each image, generated a search query (2) using each image's url as the query. For each image in C, perform a Google Image search to get the number times the images appears on the Web e_n (3). The number of returned search results determines popularity (e_n).

To create a function that normalizes e_n (4), take the sum of the logarithms of the popularity of images from all compositions, deriving average popularity ($\bar{c}$). Then double average popularity ($\bar{c}$) to estimate a maximum, and calculate the normalized image novelty ($inov$) of an element (e). This effectively sets a maximum popularity.

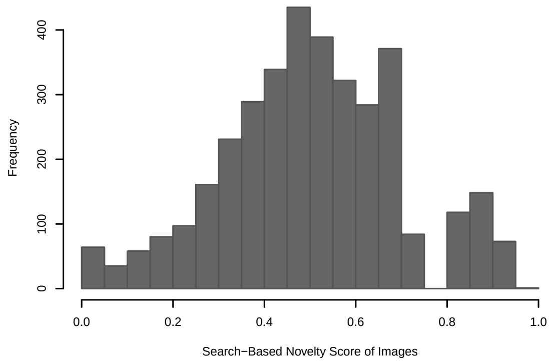

Figure 2. Distribution of the search-based image novelty ($nov_i(e)$) for 3,579 image queries.

$$\bar{c} = \frac{\sum_{\forall e \epsilon C} log(e_n)}{|e \epsilon C|} \quad (3)$$

$$inov(e) = 1 - \max(1, \frac{\log e_n}{2 \cdot \bar{c}}) \quad (4)$$

Thus, $inov(e)$ is a function for novelty that does not depend on a corpus of collected responses. $\bar{c}$ does not need to be calculated for every experiment. We found $\bar{c} = 5.02$, to get the resulting function $inov(e) = 1 - \max(1, log(e_n)/10.04)$. $inov$ will always be between 0 (not novel) to 1 (very novel). We show the histogram for $inov$ scores over C in Figure 2. Figure 1 shows example images such as the Facebook logo, which is the least novel.

FIELD STUDY

Laboratory experiments help researchers focus on factors that contribute to creativity by providing a controlled environment. In contrast, investigations in the field provide ecological validity. We conducted a field study in The Design Process: Creativity and Entrepreneurship (DPCE), an interdisciplinary undergraduate course.

DPCE students used the creativity support tool, InfoComposer [10], to author information compositions on soft innovations – new ideas formed from combining and extending existing ideas. Students searched the Internet, collecting relevant image and text bookmarks to represent ideas about their soft innovations. Students' search queries and potential information sources were *unrestricted*.

To calculate the novelty of a composition, which contains a set of images, we first calculate novelty for each of its images and then calculate a mean. We compare the prior corpus-based image novelty (Equation 1, Figure 3) and the new search-based image novelty (Equation 4, Figure 4).

Corpus-Based Image Novelty Results

Corpus-based image novelty yielded a small range of values, leaning heavily toward high novelty. Figure 3 illustrates how

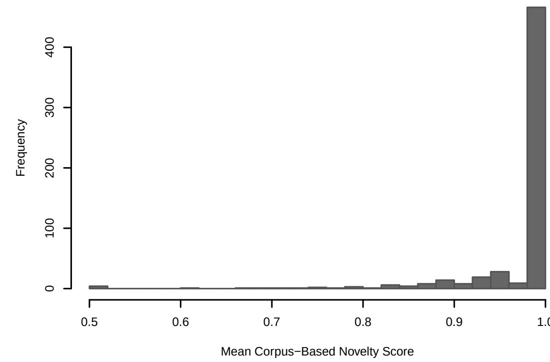

Figure 3. Distribution of corpus-based image novelty scores of 682 student authored compositions from DPCE. Even with the large number of student made compositions, the lack of overlap creates a small range of image novelty scores.

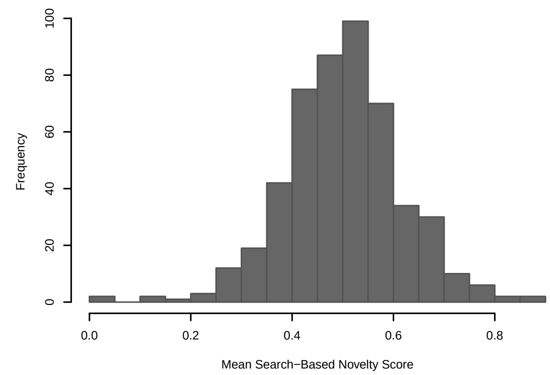

Figure 4. Distribution of search-based image novelty scores of 682 student authored composition from DPCE. Using the same images used in Figure 3, search-based novelty produces novelty scores with more granularity.

poorly the metric is suited for non-fixed queries, showing a value of 1 for most compositions. Such extreme values have do not contain enough granularity for valid comparison.

Search-Based Image Novelty Results

Search based image novelty scores provided a higher range of values. These values are what one would expect for novelty, showing a range of values. Avoiding ties in values makes this metric better suited for statistical test that require interval values, which are useful for comparing creativity support environments in field studies and laboratory experiments.

DISCUSSION

In this section, we discuss how search-based methods for measuring the novelty of elements could be applied to other kinds of media. We discuss benefits of using automated metrics to analyze products made with creativity support environments.

More Media

We have shown search-based novelty more effectively describes image novelty than corpus-based novelty when images can be collected from anywhere on the Web. While we have only implemented and tested search-based image novelty, we suspect that similar implementations would work with other kinds of media. Once one can generate queries and perform searches for an element of an arbitrary media type, then implementation of a search-based novelty score is straight forward. We envision this metric being used to assess elemental novelty for collections of various media including, audio, text, or video. For text, one could perform searches using samples of extracted text. For audio, one could search using queries from tags.

Aggregating Novelty

Justifying the methods for aggregating elemental novelty into a single score is difficult. We have used the mean of each element to represent the novelty of a collection, but we do not have a clear understanding of which aggregation would be best. As researchers, we hope to find significant differences between conditions in experiments. Transformations designed to accentuate difference could be used on these metrics, but should not be used without theoretical justification.

Elemental novelty does not consider the context of a collection, ignoring relationships between elements. If a collection includes a very popular image among moderately novel images, the overall novelty of the collection may be low. If an author creates novel combinations of ideas with common elements, they may score low on elemental novelty.

We continue to develop human rated metrics that measure holistic components of creativity [6]. While human rated metrics take longer to calculate than automated processes, they can be used in small experiments. Using crowdsourcing services can help mitigate the time any one individual has to spend to rate aspects of collections.

Implications for Design

We have shown that metrics that work well in more constrained contexts can perform poorly without fixed search queries. Big data is a resource that can be used as a baseline for the originality of content. If researchers can uncover properties that leverage publicly available web services to analyze media, they can use those processes to measure the effectiveness of tool that help people create media rich products. Combining empirical results with experiential data ensures that both the experience and the product of creativity are analyzed.

Statistical methods on computational processes for evaluating creative products help researchers analyze more data because they do not require human analysis. Computational methods can be easier to verify and repeat across research groups. Evaluation at web scale is of interest, because it involves larger sets of users, which are essential to investigating the validity of at least some creativity support environments, such as large university courses, and massively open online courses. Such large scale deployments of creativity support environments necessitate techniques that enable equally large scale automatic assessment of ideation metrics.

Conclusion

We used observations and data from a field study to show that search-based image novelty measures are more effective when the corpus of collected images yields few commonly collected images. By leveraging Google Images, we computed a baseline for image novelty to construct a useful normalized novelty metric. We presented directions for future work that extend this metric to other media types that leverage search to measure novelty.

REFERENCES

1. Arnheim, R. *Visual Thinking*. UC Press, 1972.
2. Carroll, E. A., Latulipe, C., Fung, R., and Terry, M. Creativity factor evaluation: towards a standardized survey metric for creativity support. vol. In Proc. Creativity and cognition (2009).
3. Dow, S., Fortuna, J., Schwartz, D., Altringer, B., Schwartz, D., and Klemmer, S. Prototyping dynamics: sharing multiple designs improves exploration, group rapport, and results. vol. In Proc. CHI (2011).
4. Glenberg, A. M., and Langston, W. E. Comprehension of illustrated text: Pictures help to build mental models. *Journal of Memory and Language 31*, 2 (1992), 129 – 151.
5. Kerne, A., Koh, E., Smith, S., Choi, H., Graeber, R., and Webb, A. Promoting emergence in information discovery by representing collections with composition. vol. In Proc. Creativity and cognition (2007).
6. Kerne, A., Webb, A. M., Linder, R., Smith, S. M., Moeller, J., Damaraju, S., Lupfer, N., and Qu, Y. Components of information-based ideation productivity. *submitted to ToCHI* (2013).
7. Salton, G., and Buckley, C. Term-weighting approaches in automatic text retrieval. *Information processing & management 24*, 5 (1988), 513–523.
8. Shah, J. J., Smith, S. M., and Vargas-Hernandez, N. Metrics for measuring ideation effectiveness. *Design Studies 24*, 2 (2002), 111 – 134.
9. Shah, J. J., Smith, S. M., Vargas-Hernandez, N., Gerkins, D., and Wulan, M. Empirical studies of design ideation: Alignment of design experiments with lab experiments. In *Proc ASME Conf on Design Theory and Methodology* (2003), 1–10.
10. Webb, A., Linder, R., Kerne, A., Lupfer, N., Qu, Y., Poffenberger, B., and Revia, C. Promoting reflection and interpretation in education: Curating rich bookmarks as information composition.
11. Webb, A. M., and Kerne, A. Integrating implicit structure visualization with authoring promotes ideation. In *Proc. JCDL* (2011), 203–212.

The Observable Creative Process and Reflection-on-action

Vincent Akkermans
Queen Mary, University of London
vincent.akkermans@eecs.qmul.ac.uk

Geraint A. Wiggins
Queen Mary, University of London
geraint.wiggins@eecs.qmul.ac.uk

Mark B. Sandler
Queen Mary, University of London
mark.sandler@eecs.qmul.ac.uk

ABSTRACT
The way in which creativity support tools guide creative thinking as well as the externalisation of it largely remains an unanswered question. Creative professionals should be aware of these effects so that they can better develop their practice. The proliferation of software in the creative process provides opportunities for the development of features that allow professionals to reflect on the creative process and thereby to improve this awareness. This work proposes a methodology for capturing and analysing creative processes that are mediated by creativity support tools. The captured data is described in terms of an ontology and the analysis of the data is based on a sequence-detection algorithm. The methodology is expected to lead to insights into the role that creativity support tools play in the creative process as well as to design guidelines for representations of the creative process that are useful for reflection.

Author Keywords
creativity, evaluation, innovation, art, design, user studies, ontology, observable creative process, reflection-on-practice, workflow sharing, creativity support tools.

ACM Classification Keywords
H.5.2 [INFORMATION INTERFACES AND PRESENTATION (e.g., HCI)]: User Interfaces

INTRODUCTION
In the study of the phenomenon of creativity there is much attention for its societal, social, and cognitive aspects, but arguably there is a lack of attention to how a personal creative process is brought into the world and what role software tools play in the externalisation of that process. Csikszentmihalyi [4] provides an account of the systemic properties of creativity that underlines the importance of the domain, that is the background knowledge, and field, that is the community that assesses, to creative activities and outcomes. Boden [1] emphasises that personal creativity is different and that its cognitive aspects can be studied through computational modelling. However, the externalisation of cognitive creativity, which is a necessary step for it to be assessed by others, is a process for which little theory has been developed. What are the general qualities of this process and what is the role of externalisation and reflection on the externalised in the cognitive process of creativity? Schön [11] characterised the design process, which is arguably a creative process, as a “reflective conversation with a unique and uncertain situation”. His theory of reflection-in-action provides much useful language that is directly applicable to the discussion of the observable creative process.

In many design professions, such as architecture, 3D animation, or music composition, software tools are used extensively in the daily workflow. In many cases those software tools, or creative support tools (CSTs), accompany every stage of the creative process, from making sketches to post-production. The integration of CSTs in the creative process goes much further than being an extension of the body which allows for more efficient elaboration of a mentally achieved creative result. It could be argued that CSTs are explicitly represented in creative thought processes and that they guide greatly the externalisation of the creative process. The questions of how extensive this effect is and how aware of this effect the user is remain. Given that it is the process that determines the outcome, and that therefore creative outcomes depend on innovation in the process, then surely, in professions where innovation is important the practitioners should have a good understanding of this effect.

The study of CSTs aims to understand what qualities enhance, provoke, or disrupt creativity, with the aim of being able to design better ones [9]. The field of study has a clear focus on the observable creative process and studies in the field result in a) design guidelines, such as the user being able to make variations and switch between them [7], or b) in specific designs, such as easy previews of the effect of transformations [12]. In general, these methods aim to facilitate reflection-in-action, that is to allow the user to easily explore a design problem. However, supporting reflection-on-action, which is the reflection on a completed creative process, is generally not stated as a goal. Here we extend Schön’s definition of reflection-on-action to include not only reflection on the result of action, but also reflection on the process that lead to that result.

The proliferation of software tools in the creative process presents us, not just with questions, but also with opportunities. As the creative process is mediated through software its externalisation can be logged and the captured data can be analysed. Although methods of studying user behaviour in the field of human-computer interaction through logging are not new [8], such an approach still has much to offer the study of the creative process. Not only could it provide a deeper fundamental understanding of the creative process, but it could also lead to features that allow users to more easily reflect-on-action. That is, if much of the observable creative process is captured then it can be represented to the user visually or otherwise. As argued earlier, innovation of the process is important in creative professions, and therefore awareness of the process is too. Reflection-on-action is the manner in which this awareness is cultivated and as such CSTs should include features that facilitate this. Furthermore, the degree to which CSTs support reflection-in-action should be included as a measure in their evaluation.

There is no reason why reflection-on-action assisted by representations could not be done by persons other than the original actor. In fact, the possibility of having access to information about the process of someone else might be an even more compelling reason for considering implementing reflection-on-action features in CSTs. Consider, for example, a student of composition who has a complex homework task. For the teacher being able to discuss in class the process by which the student completed the homework with the help of a visual representation could help greatly. Also, web-based communities, such as Instructables[1], where users share workflows and know-how could benefit greatly from having their creative processes automatically described and visualised. In such cases, the creativity of the community as a whole could be advanced. In a similar fashion the sharing of scientific workflows has been argued to have tremendous potential for scientific advancement [5].

Finally, it is the position taken here that a) the externalisation of the creative process should be examined more closely, b) that CSTs should explicitly support reflection-on-action, and c) that a promising strategy is to do so through log studies following the methodology described below. Currently this methodology has not been implemented and remains a proposal. It is however the intention of the authors to realise it in future work.

METHODOLOGY

Ontological Approach

The CST, viewed here as mediating between creator and artefact, can be seen as a description of the creative problem domain. Its affordances are implementations of frequently used sequences of actions and transformations and therefore embody the knowledge of a practitioner. As such, the affordances of the software tool can be described formally in an ontology that can be taken to describe the domain. These concrete affordances can then be grouped in abstract classes and organised in a class hierarchy. An example of such an affordance is the scaling of the X dimension transformation that can be applied to geometric primitives in 3D modelling tools. This scaling transformation would be grouped together with the scaling transformations for the other dimensions.

It is important to note that these affordance classes do not map directly to UI events, such as button clicks and number field input, but that these events bring into existence instances of these classes. It is possible that an analysis of events on the level of UI interaction will reveal vital information and therefore it is required that the ontology incorporates concepts to describe UI events as well as concepts to map these events to the higher level concepts of affordances.

The underlying reason for the development of this ontology is to have the ability to describe the observable creative process, which is constituted by the actions taken by the user within the CST. The order in which these actions occur is the defining feature of the process and therefore the ontology should include concepts that can describe time ordered structures of events. In addition to time ordering, there will also be a hierarchical ordering of actions for which the ontology will need to account. For example, transposing a part of a melody in a midi sequencer such as Logic consists of several UI events. Several ontologies for describing production workflows exist, such as the Music and Studio ontologies [6,10], but the concepts included are too coarse to describe the granularity of action in the processes that are of interest here.

The context in which actions occur is key to understanding them [8]. The time ordering described above is one aspect of an action's context, but equally important is the relation of an action to the artefact which is being built. If the affordances of a tool are seen as transformation on a structure, a description of an action will only be complete if it maps an instance of an affordance to the part of the structure to which it is applied and therefore the ontology will need to support such mappings.

Once the ontology has been developed the CST will need to be adapted so that the actions taken in the tool can be captured and stored.

Data Collection

At this point users should be invited to use the CST in a controlled environment. In addition to capturing the observable creative process as described above other data collection techniques should be employed as well. Video recordings of the desktop and of the user will help greatly in the interpretation of the process data. Also, from the video recordings information about the user their thoughts and experience can be obtained by joint observer and participant interpretation [8] or by user or expert reflection [3].

Analysis

The aim of the analysis of the data is to look for sequences of actions that are typical for the observed creative process as this will characterise it best and therefore give both researcher and user insight. Quantitative analysis of the data, such as the frequency of use of affordances, will undoubtedly provide insights as well, but it is expected that qualitative analysis will be more helpful. For example, a

[1] http://www.instructables.com

user will be hard pressed to know what he should improve and how from statistical descriptions alone. The collected data will have several qualities that will make qualitative analysis easier. It will be a) highly structured in terms of time ordering and hierarchy, b) described at a semantic level due to the use of an ontology, and c) highly granular as it includes all UI events.

Because of the granularity and amount of the data the detection of sequences is expected to be achievable only through computational methods. Furthermore, there are several techniques that a sequence-detection algorithm can apply to the data so that it can detect sequences that are equivalent at different levels of abstraction. This is necessary as the algorithm would otherwise likely only find very short and very few sequences that are exactly equivalent, and not any sequences that are in fact semantically equivalent. The algorithm should look for repeated occurrences of sequences employing all of the following strategies:

- *Summation*: Two or more subsequent actions of the same affordance class can be combined into one action with a functionally equivalent outcome. It is important to note that in the summation of transformations vital information might get lost. For example, the repeated nudging of an item left and right might ultimately sum up to no transformation at all. However, the repeated nudging might indicate a process of fine-tuning, which should be recognised as such.
- *Ignoring or segmenting the parameter space*: If two or more actions are of the same affordance type and the affordance class includes parameters then ignoring some or all of the parameters can lead to making the actions equivalent. Also, the parameter space of the affordance can be segmented, possibly by clustering, and the values for the parameters can be replaced by the segment its symbolic name.
- *Abstract classes*: The affordance class of an action can be replaced by one of its super classes, which has the result of making affordances that are described by the ontology as being similar equivalent.
- *Grammars*: Rewrite rules can be used to define equivalent sequences of patterns [8]. It allows for domain specific knowledge to be incorporated into the sequence-detection algorithm.

The development of the sequence-detection algorithm and rewrite rules should be inspired by the data collected from the aforementioned joint participant and observer interpretation. Similarly, the results of the algorithm should be consistent with this data.

It is expected that findings at this point will lead to improvements of the analysis method in an iterative manner. Hopefully the results of the analysis will lead to an enhanced understanding of the creative process as it is externalised through software tools. Also, any findings will inform how to visualise or otherwise represent the observed creative process to users.

Experiment

As mentioned in the introduction a key reason for proposing the methodology described is to give the user the ability to reflect on their creative process as it is the contention that it will help them to develop their creativity over time. With the insights gained from the data collected and analysed visual representations of that data can be developed. The hypothesis is that there will be a measurable difference in user behaviour when they have the ability to explicitly reflect on their own creative process with the help of visualisations.

To be able to measure this hypothesised change a first set of users, Group A, will need to participate in several sessions. After each session the users will be asked to reflect on their own creative process with the help of visualisations. A second set of users, Group B, will go through the same sessions, but will not get the opportunity to reflect. An analysis method will need to be developed that can characterise processes at a high level and can show that there are differences between the processes from the later sessions of Group A and Group B.

In addition to looking at the differences between processes, an ethnomethodological approach can be taken to look at how users interact with the visualisations. Finally, through questionnaires or interviews users could be asked what their opinion or experience of interacting with the visualisation was. Of special interest is whether users will have become more aware of what their process has been; that is whether it has helped them reflect-on-action.

DISCUSSION

There is an important difference between the methodology used in usability studies that use data collected through logging and the methodology proposed here. Whereas in a usability study one will tend to look for problems with the interface, these problems are not the subject of investigation here. It is the process itself and its intrinsic qualities that are of import.

The ontological approach to the description of creative processes brings with it some disadvantages that are worth mentioning. First of all, whoever develops the ontology needs considerable knowledge of how the CST works and of the creative problem domain the CST is designed to address. Secondly, in ontological work there are often multiple valid but different ways of describing a concept and it is therefore difficult to evaluate the developed ontology [2]. Lastly, it is unclear how much of the ontological work can be reused for other CSTs.

As mentioned in the introduction, the proposed methodology is mainly concerned with CSTs that are used by professionals. The majority of these tools are commercial and closed sourced software and are therefore unlikely candidates to be picked as focus of this methodology. If an open source CST is picked it is however important that a) it is comprehensive in the sense that it has a similar feature set to its commercial counterparts, and b) it has a large enough user base to find participants for the experiments. If a CST is not comprehensive it cannot be used from start to end of the creative process and as such the captured data will be incomplete either because the process cannot be finished or because for some parts of the process other tools were used.

Although to some degree the difficulty of adapting a CST depends on the complexity and source code quality of the software, for any reasonably comprehensive CST this will be a time consuming task. It seems a good strategy to implement as much of the supporting functionality, such as storage and analysis of data, as possible in a service oriented architecture and thereby keeping the adaptations to the CST to a minimum.

The proposed methodology is a risky endeavour because of several extensive preparatory steps that will need to be taken, such as the development of the ontology and the adaptation of the CST. On the other hand, it could result in valuable insights into creative processes as well as useful new functionality for users.

CONCLUSION

In many professions CSTs are used extensively in all stages of the creative process. However, the role CSTs play in the externalisation of the creative process is not well known. Similarly, CSTs guide creative thinking and therefore both enable and limit it. For creative professionals it is important to develop their creative processes and they should be aware of what they are and how CSTs influence them. CSTs should therefore include features to make this reflection-**on**-action possible.

Furthermore, a methodology is proposed for capturing and analysing observable creative processes. It is based on the development of an ontology, which describes the affordances of the CST and the domain of creativity. The ontology should also include concepts that describe events and their chronological and hierarchical ordering, UI events, and links to part of the structure of the artifact that transformations are applied to.

Also, an abstract sequence-detection algorithm to find repeated occurrences of sequences of actions is described. It can employ techniques such as summation, ignoring or segmenting the parameter space, substitution by abstract classes, and rewrite rules to find sequences of varying equivalence.

Findings from the development of this analysis method are expected to lead to insights into the creative process and to inspire ways of representing it to the user. Finally, these representations and their usefulness will have to be evaluated in their own right.

ACKNOWLEDGMENTS

Support from the Media and Arts Technology Programme: An RCUK Doctoral Training Centre in the Digital Economy.

REFERENCES

1. Boden, M.A. *The creative mind: myths and mechanisms*. Routledge, London, England, 2004.
2. Brewster, C., Alani, H., Dasmahapatra, S., and Wilks, Y. Data driven ontology evaluation. *Proceedings of the International Conference on Language Resources and Evaluation*, (2004), 24–30.
3. Carroll, E.A. and Latulipe, C. Triangulating the personal creative experience: self-report, external judgments, and physiology. *Proceedings of Graphics Interface 2012*, (2012), 53–60.
4. Csikzentmihalyi, M. *Creativity: Flow and the Psychology of Discovery and Invention*. HarperPerennial, 1997.
5. De Roure, D., Goble, C., and Stevens, R. Designing the myExperiment Virtual Research Environment for the Social Sharing of Workflows. *Proceedings of the Third IEEE International Conference on e-Science and Grid Computing*, (2007), 603–610.
6. Fazekas, G. and Sandler, M.B. The Studio Ontology Framework. *Proceedings of the 12th International Society for Music Information Retrieval Conference*, (2011), 471–476.
7. Hewett, T., Czerwinski, M., Terry, M., et al. Creativity support tool evaluation methods and metrics. *Creativity Support Tools, A workshop sponsored by the National Science Foundation*, (2005), 10–24.
8. Hilbert, D.M. and Redmiles, D.F. Extracting usability information from user interface events. *ACM Computing Surveys 32*, 4 (2000), 384–421.
9. Kerne, A., Webb, A.M., Latulipe, C., et al. Evaluation Methods for Creativity Support Environments. *ecologylab.net*, 2012. http://ecologylab.net/workshops/creativity/CHI_CSE_Workshop.pdf.
10. Raimond, Y., Abdallah, S., Sandler, M.B., and Giasson, F. The Music Ontology. *Proceedings of the 8th International Conference on Music Information Retrieval*, (2007), 417–422.
11. Schon, D.A. *The Reflective Practitioner: How Professionals Think In Action*. Ashgate Publishing Limited, Farnham, England, 1991.
12. Terry, M. and Mynatt, E.D. Recognizing creative needs in user interface design. *Proceedings of the 4th conference on Creativity & Cognition*, (2002), 38–44.

Techniques for Evaluating Novice-Oriented Creativity Support Tools

Nicholas Davis, Alexander Zook, Mark Riedl
School of Interactive Computing
Georgia Institute of Technology
{ndavis35; a.zook; riedl}@gatech.edu

Friedrich Kirschner
Hochschule fuer Schauspielkunst Ernst-Busch
Berlin, Germany
f.kirschner@hfs-berlin.de

Michael Nitsche
School of Literature, Media, and Communication
Georgia Institute of Technology
michael.nitsche@gatech.edu

ABSTRACT

We present evaluation techniques for a novice-oriented creativity support tool in the domain of digital filmmaking. Novices need help *executing* tasks as well as *knowing* which tasks are appropriate and the implications their decisions have for their creative product. With our tool, we focus on supporting critical domain knowledge that can enable novices to make meaningful creative contributions. In film, domain knowledge includes cinematographic and editing rules and conventions. Our creativity support tool (CST) provides feedback when novices violate these norms. We use two approaches to evaluate the results: (1) Expert Consensual Assessment, and (2) Individual-Group Consensus, which is a specialized technique we developed to overcome the limitations of the Consensual Assessment Technique for this application. We discuss these two techniques and their domains of application. Finally, we call for further research into evaluation techniques for CSTs sensitive to the task evaluated (e.g. execution vs. knowledge support) and relevant domain (e.g. machinima).

Author Keywords
Creativity Support Tools; Evaluation; Cognitive Science

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Human Factors; Design; Measurement.

INTRODUCTION

In this paper we investigate evaluation techniques for tools to support creative digital filmmaking. We study machinima, which is a new form of digital filmmaking that leverages the real time graphic rendering capabilities of video game engines to create polished animations. Machinima films began as a recording of scripted video game characters with audio overlay. However, the tools of the trade have expanded beyond the initial confines of early video game engines to introduce much more complexity and nuance and include control of lighting, set and character design, and cinematography. This technology adaptation has opened a door to individuals with no experience in animation or filmmaking to create professional looking animated films. The proliferation and open nature of machinima tools have introduced many new avenues for creative expression and significantly lowered the barrier to entry for digital filmmaking. These tools have empowered individuals to creatively express themselves in ways that were prohibitively expensive a decade ago. However, despite how powerful the tools have become, many elements of the filmmaking craft remain unknown to novices.

There is a filmic language that denotes standards, conventions, and general cinematographic rules that experts gradually learn through education and experience [1]. Although experts often violate these rules for stylistic purposes, novices unknowingly violate these rules, which can interrupt the visual continuity between shots, temporal rhythm, and spatial orientation of a scene.

Our earlier study [3] found novices routinely violate many of these filmmaking conventions, and the creativity of their products suffers. Although the graphics of their film may look sophisticated, creative decisions that unintentionally violate cinematography and editing conventions distract the viewer from the story. The study reveals novices require creativity support in two ways: (1) *executing* known creative tasks and (2) *knowing* the norms and values of a creative domain. How can we help novices avoid errors associated with execution and knowledge to produce higher quality creative content?

Lubart [9] enumerates four ways computers can support creativity. A computer *nanny* provides tools to schedule and maintain creative activities. A computer *pen-pal* supports collaboration within teams. Computer *coaches* stimulate creative thinking by suggesting creative activity based on expert knowledge in the domain. Computer *colleagues* may meaningfully contribute to tasks so a human-computer team becomes a contributor to a domain.

Novices require creativity support tools (CSTs) that provide guidance like a coach and also perform skilled operations like an expert colleague that simplifies the process of creation. Execution of a known creative goal is not sufficient support for novices. Rather, coaching must support novices acquiring the deep domain knowledge experts possess. Lacking this knowledge, novices are unable to understand the implications of their decisions for achieving creative outcomes. Novices are often unable to contextualize their creative goals within a new domain without this knowledge.

The video game engine in the machinima tool performs the expert task of animating content, but users also need a coach that provides feedback about their decisions. To address this need we designed an intelligent creativity support system that analyzes camera placement to determine if the user has violated cinematographic rules, such as those listed below.

- **180 Degree Rule:** The line of action is an imaginary line created between two individuals engaging in a conversation. Once a camera is placed on one side of the line of action, subsequent camera angles should remain on that side of the line of action to prevent the characters from changing their perceived location [7].
- **30 Degree Rule:** The degree of change between two sequential camera angles should be greater than 30 degrees to reduce jumpy or jittery cinematography [5].
- **Cutting on Action**: Camera angles should switch during an action being performed, rather than immediately before or after the action. This creates the illusion of fluid movement [2].
- **Pacing**: Shots should change at a regular frequency and abrupt changes should only occur during highly emotional or dramatic moments [8].

Unlike a simple error-correction approach our tool aims to help inform novice decisions. Feedback provides information and explanations of violations to inform users without constraining them to follow the norms of cinematic practice. Existing machinima tools support novice *execution* needs; our approach targets novice *knowledge* needs. We balance the creative freedom of the user with the need for domain-specific knowledge.

In the remainder of the paper we first describe an experiment to evaluate our machinima CST. We then present two techniques to evaluate the novice films created in our study to understand the impact of our CST. This analysis addresses successes and limitations of the approaches. We conclude by discussing the broader application of one of the techniques and call for further analysis of CST evaluation techniques specific to evaluation tasks – such as execution or knowledge support – and evaluation domains – such as machinima creation.

Experiment Design

We will only briefly outline the experiment design here since this paper focuses on CST evaluation techniques. Our experiment investigated the effect of offloading expert knowledge about rules onto digital filmmaking CSTs [4]. We hypothesized that providing knowledge about the filmic language would help users make more informed decisions as they relate to rules, manifested in fewer violations of those rules. Our study found this to be the case, see [3] for more detailed results.

Evaluating systems that support novices is difficult because they lack baseline execution skill and knowledge in the application domain. Novices may have difficulty executing a desired task using the interface and knowing appropriate domain content to use. To minimize execution problems, we isolated and simplified the required functions for the task. Participants were only responsible for selecting the specific angle and timing for a shot from a list of pre-determined camera angles. This limited the learning curve for approaching the software and focused our evaluation on novice cinematic knowledge support.

We recruited 20 participants for this study. Participants were split into two groups that both engaged in a constrained creative activity using the Xtranormal machinima software (www.xtranomral.com). We selected Xtranormal because it is a popular and freely available machinima creation tool. A pre-scripted scene and pre-defined set of expert selected camera angles were provided to the participants to reduce the amount of training and tool expertise required for the task. The task was to find the best editing and cinematography for the scene with the given materials, which meant selecting the appropriate camera angle and deciding when it should be used in the scene. Users could insert or remove any number of camera angles from the list during the 40 minutes they were provided for the task. The control group (n=10) was unaided. The experimental group (n=10) was able to press an 'Analyze' button that prompted the system to analyze their camera selections and provided feedback about any rules they violated.

The program that analyzed the films was a Wizard of Oz (WOZ) system in which a human expert evaluated the user's decisions and sent standardized feedback to the user based on the rules that they violated. Users were able to request feedback on their current selections from the system whenever they desired. Feedback was provided through an IRC chat channel that appeared next to the Xtranomral interface. The participants thought the feedback came from an automated system.

The 'wizard' watched each film in real time and noted all current errors. The wizard's assistant entered these errors into an interface using template text explanations. When participants pressed the 'Analyze' button on the WOZ interface, the assistant sent a message that contained a list of the user's rule violations as well as explanations describing those rules, similar to the explanations above.

Evaluation Techniques

We evaluated our CST by comparing the resulting novice films using both a standard expert evaluation methodology and a specialized method developed for our domain. Evaluation focused on novice errors as our previous study found these are required to recognize creative success in the machinima domain [3]. Below we describe the two consensus-based evaluation methods we used on our CST and their strengths and limitations.

Expert Consensual Assessment

One approach to evaluate the effectiveness of our CST is to use the Consensual Assessment Technique (CAT) [6] where a panel of experts rates the users' products. We attempted this evaluation by recruiting three film experts to independently rate the general and technical quality of each of the film clips. However, these rating had a low inter-rater reliability and were therefore discarded as invalid. There are several factors that may have contributed to this result.

One explanation could be that CAT may be less appropriate when the creative products are extremely similar. Since we studied novices who were not familiar with the machinima tool or creating films, we simplified their contribution in way that minimized the necessary skill. This also changed the degree of creative freedom that participants had to create diverse products. The experiment was designed to constrain the creative task to selecting the angle and timing of camera placement from a pre-defined list of camera angles. The set of film clips created were approximately 1-minute long and appeared very similar to each other visually. The camera angles had significant variations, but the overall visual similarity of the scenes may have reduced the evaluators' sensitivity. It may have been hard to make meaningful comparisons between the clips.

A second explanation could be that evaluating film is fundamentally different than evaluating static images or products (as is more traditional in CAT) because they cannot be placed next to each other for comparison. Timing effects may come into play since the product is experienced through time. To mitigate this, we encouraged evaluators to take notes of important information. However, sequential evaluations can skew expert ratings due to anchoring on early aspects of a film clip or earlier clips in a series [10].

A third factor is that the time and focus required for a series of ratings may lead to evaluator fatigue. Evaluating films typically takes more time than evaluating static images as each film has to be watched in its entirety. More rapid fatiguing makes it difficult to reach consistent evaluations from a set of experts when examining more than a few movie clips.

Individual-Group Consensus Evaluation

To bypass the limitations of CAT in our domain we developed a method of gathering individual evaluations followed by a group consensus process. Three researchers who helped design the experiment analyzed each film to determine rule violations. It was important to select individuals familiar with the design because they would be able to effectively detect rule violations. The analysts were familiar with the pre-determined camera list and had a sense of which shot combinations typically constitute an error.

Each analyst watched all the clips (in random order) independently and noted any errors. There were four rules in total, and each shot could have multiple errors. The analysts watched each individual shot and noted the errors they detected individually. Using analysts familiar with the domain, aware of experimental constraints, and taking detailed notes on error timing helped circumvent the limitations of applying CAT discussed above.

The analysts aggregated their error judgments for a combined evaluation. Multiple possibilities for combining these ratings exist. One option is to use a majority rules voting system for the error data for each shot. In a majority rules approach, whichever decision had at least two-thirds support would be selected. This approach is faster because it is accomplished without viewing the film clips again and occurs without much discussion. However, it runs the risk of overlooking crucial data due to a shallow analysis of differences in interpretation among raters. When employing CAT we found experts would agree on qualitative classes of errors, but differ in the specific points they found problematic. A majority rules approach suffers this same limitation.

We decided to instead have all three analysts view all the clips in their entirety a second time as a group. Each clip was analyzed fresh and the analysts came to a consensus agreement on each clip. The group decision was then compared to each of the individual analysts' error evaluation. If there was a discrepancy between the group decision and any of the individual analysts' error evaluation, a further investigation was pursued. The analyst(s) whose error evaluation contradicted the group decision was asked to explain why he or she made that decision, and if s/he still felt the same way. The group would then discuss this information and decide which decision to maintain.

We preferred this approach because detecting errors was difficult given that each 1-minute scene contained between 5-15 shots and there were approximately 30 clips. It is easy for one person to overlook an error. Introducing multiple viewers reduced the likelihood of missing an error. Additionally, the individual data helped to highlight potentially inaccurate decisions the group came to. There were more redundancies in this approach, which is preferable because it presented more situations to challenge and rectify the evaluation decisions. We found the group consensus approach helped detect many overlooked errors and correct for differing interpretations among evaluators, particularly in regards to pacing rule violations.

CONCLUSIONS

Novices often lack both the skills to *execute* a desired creative goal and the *knowledge* to be aware of whether their execution is within the norms and values of a domain. Evaluating creativity support tools that are targeted for supporting knowledge and reducing fundamental errors may require specialized evaluation procedures. Our CST was designed to support the informed exploration of camera angles and provide feedback to the user about when their decisions violated important cinematic norms. This type of feedback encodes expert knowledge into the CST and enables novices to achieve better creative outcomes. We argued that testing the effectiveness of this type of support is best measured using an individual-group consensus approach with detailed product analysis.

Several factors may have contributed to invalid CAT judgments, including the similarity of products, the temporal nature of film, and evaluator fatigue. We circumvented these limitations through a modified technique that retained expert evaluation and consensus while introducing greater awareness of task constraints, a round of detailed and recorded evaluation, and a consensus approach focused on aligning group and individual assessments. The individual-group consensus process adds much needed redundancy to creativity evaluation with more careful accounting of disagreements.

The technique we describe is applicable to a wide variety of creative tasks, but is particularly useful when examining temporally extended or “large-scale” artifacts. CST evaluation often requires detailed information on the effects of a CST intervention on a given product, rather than only a holistic judgment of the resulting changes in user creativity. In our case detailed judgments of knowledge-related errors were required; other domains likely require other metrics. An individual-group consensus approach centered on these detailed evaluation points can bypass cognitive biases (e.g. anchoring) and limitations (e.g. fatigue) that hinder the use of holistic judgment in these domains. A further benefit of this method is providing detailed information to inform the further development of CSTs, documenting the relative magnitude of different aspects of creativity in a target domain. Ultimately, we hope this work highlights the need to develop CST evaluation techniques that are sensitive to the evaluated task (execution vs. knowledge support) and evaluation domain (e.g. machinima).

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. IIS-1002768. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

1. Arijon, D. (1976). Grammar of the film language.
2. Beller, H. *Handbuch der Filmmontage. Praxis und Prinzipien des Filmschnitts.* Tr Verlagsunion , München, 1999.
3. Davis, N., Zook, A., O’Neill, B.O., Grosz, A., Headrick, B., Nitsche, M., Riedl, M. Creativity Support for Novice Digital Filmmaking. To Appear in *Proceedings of Human Computer Interaction (CHI) 2013.*
4. Davis, N., Li, B., O’Neill, B., Nitsche, M., and Riedl, M. Distributed Creative Cognition In Digital Filmmaking. In *Proc. Creativity & Cognition 2011*, ACM Press (2011), 207–216.
5. Hayward, S. *Cinema Studies: The Key Concepts*. Routledge, London, 2000
6. Hennessey, B. A., Amabile, T. M., & Mueller, J. S. (1999). Consensual assessment. *Encyclopedia of creativity*, *1*, 346–359.
7. Katz, S. D. *Film Directing Cinematic Motion: A Workshop for Staging Scenes*. Michael Wiese Productions, Studio City, 2004.
8. Kindem, K. and Musburger, R. *Introduction to Media Production: The Path to Digital Media Production.* Focal Press, Burlington, 2009.
9. Lubart, T. How can computers be partners in the creative process: Classification and commentary on the Special Issue. *International Journal of Human-Computer Studies* 63, 4-5 (2005), 365–369.
10. Tversky, A. and Kahneman, D. (1974). "Judgment under uncertainty: Heuristics and biases". *Science*, 185, 1124–1130.

Design Creativity: Using Pareto Analysis and Genetic Algorithms to Generate and Evaluate Design Alternatives

Erin Bradner
Autodesk Inc.
One Market, Suite 200
San Francisco, CA, 94105 USA
erin.bradner@autodesk.com

Mark Davis
Autodesk Inc.
One Market, Suite 200
San Francisco, CA, 94105 USA
mark.davis@autodesk.com

ABSTRACT

When presented with a design problem, engineers, architects and visual-effects artists routinely generate multiple, competing design alternatives. Generating alternatives enables designers to explore the creative space along multiple parallel and orthogonal dimensions. This paper examines a new design practice called *design optioneering*, which uses parametric modeling, Pareto analysis and genetic algorithms to simultaneously generate and evaluate large sets of design alternatives. We offer examples of built-in evaluation methods within design optioneering tools. We argue that a tool's ability to automatically evaluate design solutions against users' predefined fitness criteria, greatly filtering a large set of alternatives generated, may be the most relevant and effective evaluation of creativity support for this class of tool.

Author Keywords

evaluation, innovation, design, creativity support tools

ACM Classification Keywords

I.5.2 Design Methodology

General Terms

Human Factors, Design, Theory

INTRODUCTION

Design optioneering tools are creation tools that use parametric modeling, performance simulation and design optimization to systematically generate and evaluate design alternatives [7,8]. Design optioneering is a departure from customary architecture and engineering practice. Customarily, architects design a relatively small set of design alternatives which represent specific points in a multi-dimensional design space [5]. In architecture, this small set of design alternatives may be communicated in the form of two or three laser cut scale models or a few dozen photo-realistic visualizations. Even with the support of state-of-the-art computer-aided design tools, individual designs are iterated relatively slowly and with considerable design effort.

Conversely, architects and engineers using design optioneering practices generate orders of magnitude more design alternatives by specifying design objectives in the form of design parameters and parameter ranges. They use genetic algorithms to automatically and iteratively generate large sets of design alternatives [7]. The designs that best fit the architects or engineer's predefined parameters "survive" multiple "generations," i.e. iterations, and mutate to spawn successive generations of uniquely new designs.

Contrasting with traditional design practice, *optioneered* designs are computed parametrically and bred algorithmically. The numerous design alternatives that are produced are often represented by a multi-dimensional plot of the solution set and might be coupled with a matrix of thumbnails of rendered designs, as in Figure 1. Researchers investigating the approach argue that design optioneering is unique in that it "enables designers to more efficiently and with more certainty explore more complex and tightly coupled design solution spaces." [ibid]

DESIGN OPTIMIZATION AS A CSE

Design optioneering tool sets are an under-researched class of Creativity Support Environments (CSEs). To the extent that design optioneering tools empower users to be "not only more productive but also more innovative" [10] they satisfy the broad definition of CSEs (a.k.a. CSTs) put forth by Shneiderman et al in their 2006 NSF workshop report. As a tool for creative thinking, optioneering tools support exploration, experimentation, and by definition they generate numerous combinations of multiple design parameters. Gerber [7] articulates the following six features of the design optioneering approach:

1. rapid design exploration
2. increased integration between disparate design and engineering domains
3. rapid visualization of the cause and effect of design decisions
4. quantified understanding of multi-objective tradeoffs
5. reduced design decision latency
6. transparency and automation of the generation and evaluation of the design solution spaces.

CHI'13, April 27 – May 2, 2013, Paris, France.

Gerber's features map closely to the design principles for CSEs in [10]. The overlapping goals include those of maximizing exploration, increasing collaboration, emphasizing iteration, and supporting evaluation.

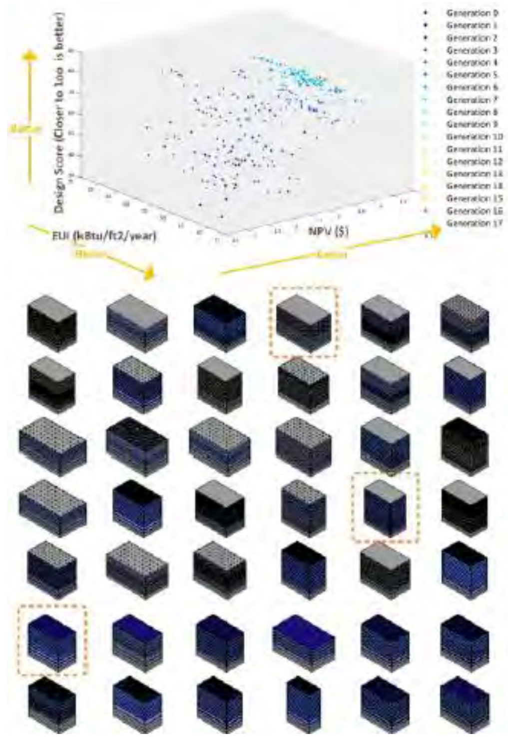

Figure 1. Sample Design Optioneering Results from [7]

DESIGN OPTIMIZATION AND EVALUATION

Evaluation of CSEs is an active area of research, motivated in part by the dearth of qualitative and quantitative models describing creativity. Carroll et al. [2] have developed, and statistically validated, six orthogonal factors to describe CSEs. These include *Results worth Effort*, *Expressiveness*, *Exploration*, *Immersion*, *Enjoyment* and *Collaboration*. The factors of creativity that are most relevant to design optioneering tools are *collaboration* and *exploration*.

Design optioneering is a creative task, with the task being how to generate and explore a set of designs that optimize against multiple, competing constraints. In building design it is commonly necessary to simultaneously consider multiple complex objectives including site utilization, structural design, building form, energy use, and construction and operation costs. The design task draws on expertise in multiple domains including structural engineering, energy simulation, architecture and finance. Improving the integration of multiple expert domains – i.e. improving collaboration – is an explicit goal of optioneering [4,6,7]. Although no quantitative measures of collaboration exist, design optioneering researchers consistently cite increased collaboration among design disciplines as a primary benefit to the approach.

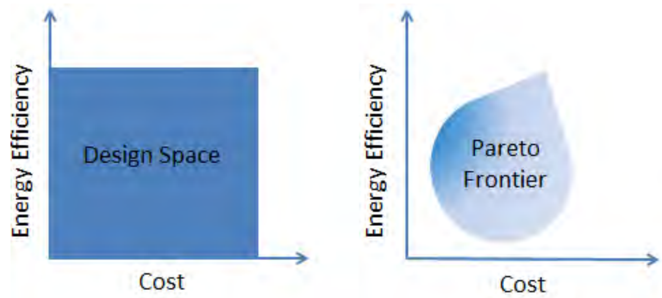

Figure 2. Design Space and Pareto Frontier (stylized)

The set of design permutations for any given design task is vast. Flager et al calculate that for the simplistic one room, steel-frame building used in their energy and structural optimization study there were $55x10^6$ possible solutions [6]. Cognitive limitations prevent users from discerning meaningful differences across solutions in sets of this size. Furthermore, time constraints prevent users from exhaustively exploring the set for top performing solutions. To address these limitations, Pareto analysis is used to winnow the set of all solutions down to a subset that is reasonably efficient across multiple objectives [7]. Figure 2 shows a stylized plot of all possible design alternatives in a design space (left) and the reduced set of optimized alternatives as computed by a Pareto analysis (right). The Pareto set identifies a subset of equally good solutions from the set of all solutions. For each member of the Pareto set there exists no other solution that is equal or better with respect to all other parameters. As an evaluation tool for creativity, the Pareto set delineates the most interesting and optimized solutions from the set of all possible solutions [9]. Calculating the Pareto set allows users to focus their design exploration among the most promising solutions.

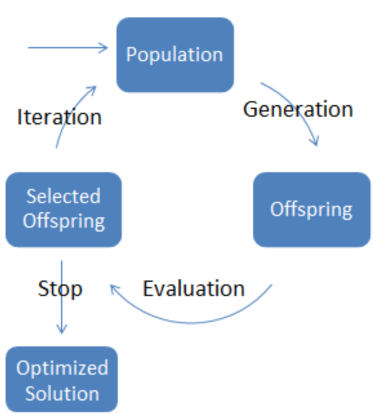

Figure 3. Typical Architecture of a Genetic Algorithm

In addition to Pareto analysis, genetic algorithms are also used to evaluate designs in design optioneering. These algorithms generate, select and breed the solution space. Users control variables such as the population size, number of iterations, probability of mutation, and stopping criteria. Figure 3 shows the typical architecture of a genetic algorithm. Genetic algorithms mimic natural selection by introducing random variation into the design process. Design researchers deliberately choose genetic algorithms over alternative optimization algorithms for their ability to evaluate discontinuous and "noisy" design spaces [6]. Analogous to natural selection, designs produced by genetic algorithms are systematically evaluated against constraints in the design environment. The most adaptive designs, i.e. the most creative designs, survive.

CONCLUSION

Pareto analysis and genetic algorithms are two built-in evaluation techniques guiding the creative output of optioneering tools. It can be argued that these evaluation techniques increase the creative capacity of the designer by generating a large number of informed guesses at desirable designs. Using design optioneering methods researchers have computationally generated and analyzed over 5,000 design alternatives in 32 hours [6] where conventional methods require 12 weeks to produce three alternatives [5]. A large set of analytically tested designs is preferable to the small set of solutions produced in the typical design project. Design theory indicates that the exploration of a large number of design alternatives is essential to successful design [3]. Software features that automatically evaluate the fitness of designs produced by the software support the creative process by focusing designers' attention on the most relevant and potentially innovative designs.

We propose that evaluation methods might best be *designed in* to CSE tools themselves. Since the creative space of design alternatives is conceptually infinite, CSEs in engineering and architecture must include evaluation support within the software that is used to generate designs. Continuous evaluation both narrows focus and expands creativity; Pareto analysis focuses the set of design alternatives to the most optimal solutions, while stochastic methods like genetic algorithms introduce randomness to expand the space.

Admittedly, this method of evaluation is lacking external validity. Shneiderman and Plaisant [11] compel us to consider that one of the most convincing measures of creativity support is a tool's ability to enable expert users to achieve their professional goals. It may be the case that the best evaluation of design optioneering tools is done by the professional consuming the designs – she is the final arbiter of the creativeness of the solution, relative to the design problem. The degree of support that a design optioneering tool provides for the creative process may be best judged by the designer's willingness to anoint one design, among many, as the most creative.

REFERENCES

1. Agrawal, G., Lewis, K., Chung, K., Huang, C., Parashar, S. and Bloebaum, C. Intuitive Visualization of Pareto Frontier for Multi- Objective Optimization in n-Dimensional Performance Space. Technical Report, University at Buffalo - State University of New York (2004).
2. Carroll, E. A., Latulipe, C., Fung, R. and Terry, M. Creativity factor evaluation: towards a standardized survey metric for creativity support. *Proc. ACM conference on Creativity and Cognition* ACM Press (2009).
3. Eastman, C., McCracken, W. M. and Newstetter, W. C. Design Knowing and Learning: Cognition in Design Education. Elsevier Ltd. (2001).
4. Fialho, A., Hamadi, Y. and Schoenauer, M. Optimizing architectural and structural aspects of buildings towards higher energy efficiency. *Proc. 13th Annual Conference on Genetic and Evolutionary Computation*. ACM Press (2011), 727-732.
5. Flager, F. and Haymaker, J. A Comparison of Multidisciplinary Design, Analysis and Optimization Processes in the Building Construction and Aerospace Industries. *Proc. of the 24th International Conference on Information Technology in Construction* (2007).
6. Flager F, Welle B, Bansal P, Soremekun G, Haymaker J *Multidisciplinary process integration and design optimization of a classroom building*. Journal of Information Technology in Construction (ITcon), Vol. 14, (2009), 595-612.
7. Gerber, D. J., Lin, S.-H., Pan, B. and Solmaz, A. S. Design optioneering: multi-disciplinary design optimization through parameterization, domain integration and automation of a genetic algorithm. *Proc. Symposium on Simulation for Architecture and Urban Design 2012*. Society for Computer Simulation International (2012).
8. Holzer, D., Hough, R. and Burry, M. Parametric Design and Structural Optimisation for Early Design Exploration. International Journal of Architectural Computing, Vol. 5, 4 (2007), 625-643.
9. Pareto, V. *Manual of Political Economy*. A.M. Kelley, New York, NY, 1971.
10. Shneiderman, B., Hewett, T., Fischer, G., Jennings, P. et al. *Creativity Support Tools: Report from a US National Science Foundation Sponsored Workshop*. International Journal of Human Computer Interaction, 20, 2 (2006), 61-77.
11. Shneiderman, B. and Plaisant, C. Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies. *Proc. of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*. ACM Press (2006).

Intelligent Creativity Support

Winslow Burleson
School of Computing Informatics and Decision Systems Engineering
699 S. Mill Avenue
Tempe, AZ
Winslow.Burleson@asu.edu

Kasia Muldner
School of Computing Informatics and Decision Systems Engineering
699 S. Mill Avenue
Tempe, AZ
Katarzyna.Muldner@asu.edu

ABSTRACT
We present an argument for the advancement of Intelligent Creativity Support (ICS) systems as an integrating framework for Intelligent Tutoring Systems, Affective Computing, and Creativity Support Tools, in a manner that closely aligns each of these technologies and research agendas with the componential model of creativity, i.e. domain relevant expertise, intrinsic motivation, and creative thinking style, respectively. We will also present strategies for the development and evaluation for student models for the just-in-time assessment of creativity.

Author Keywords
Creativity; adaptive support; affect.

INTRODUCTION

While there are over one hundred definitions of creativity [3], there is consensus that it entails a product, idea, or process that is *novel* and *useful* [26]. Creativity is at the core of all societal advancements. It is also present “not only when great historical works are born but also whenever a person imagines, combines, alters, and creates something new, no matter how small” [44].

Creativity has been described as the most vital economic resource of our time [18, 21] and the U.S. Council on Competitiveness has indicated that it will be the top factor determining America’s success in the 21st century [34, 45]. Thus, understanding how to foster creativity skills is a crucial societal goal [40]. U.S. universities, colleges, and K-12 school systems can play a fundamental role in producing an innovative and creative workforce, by helping students develop such skills [34, 41, 45]. Indeed, the 21st Century Skills initiative [39] and Common Core Standards [31] call for teaching creativity, innovation, and deep problem-solving abilities.

Unfortunately, various challenges have hindered the adoption of creativity instruction and practices in traditional classrooms [27]. For one, few teachers have been trained in how to teach creativity [25]. More importantly, classroom settings do not enable teachers to provide the individualized support needed for effective creativity facilitation. In particular, while personalized instruction has tremendous potential to improve student learning [10, 23], affect (motivation and emotion) [23, 32], and metacognitive skills [7], providing a human tutor for each student is simply not practical. Given these challenges, most of the work thus far reflects anecdotal, descriptive data [24, 34-36], although some exceptions exist [8, 9, 17].

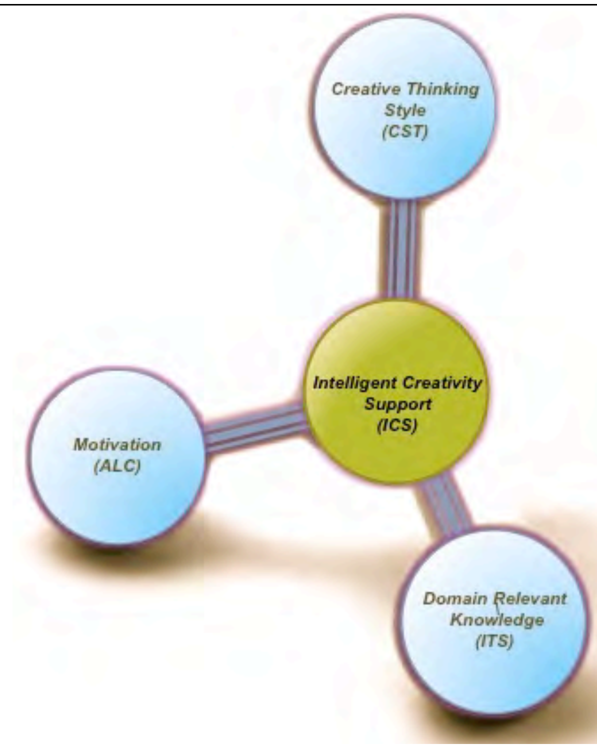

Figure 1. Advancing a new class of cyberlearning technologies, Intelligent Creativity Support (ICS) will integrate personalized support with Amabile’s componential model of creativity. ICS will combine Intelligent Tutoring Systems (ITS) to increase domain relevant knowledge; Affective Learning Companions (ALC) to foster motivation; and Creativity Support Tools (CST) to advance creative thinking styles.

Since Intelligent Tutoring Systems (ITS) can provide large-scale instruction that continuously adapts to learners' needs [1, 5, 22, 37, 43], they present a unique opportunity to address issues associated with teaching creativity. ITS have already successfully improved domain learning by tracking students' problem-solving progress, providing tailored help and feedback, and selecting appropriate problems [38, 43]. However, ITS have also been criticized for over-constraining student problem solving and over-emphasizing shallow procedural knowledge, and therefore not properly addressing 21st century higher-order skills like critical thinking and creativity [39].

We will present strategies for the design of Intelligent Creativity Support (ICS) to foster student creativity during STEM activities. The ICS framework is situated within the validated and broadly adopted componential model of creativity [2]. Amabile's model highlights three factors within an individual that are needed for creativity: domain knowledge, motivation, and creative thinking styles. Moreover, Amabile and others have demonstrated that positive affect contributes to creative problem solving [19, 20], leading to increased intrinsic motivation, deeper exploration, and more appropriate outcomes or solutions. We will discuss how ICS can integrate and leverage traditionally isolated technological components that are critical to advancing a student's creative capacity (see Figure 1): (1) domain relevant knowledge supported by ITS; (2) affect (motivation and emotion) fostered by Affective Learning Companions (ALC); and (3) creative thinking skills scaffolded by Creativity Support Tools (CST). The ICS design can also implicitly account for external factors that influence creativity, such as evaluation and time pressure [2, 3]. The overall goal of the ICS strategy is to extend traditional ITS instruction with personalized affective support and metacognitive creativity training, to improve creativity and learning outcomes.

STUDENT MODELS FOR JUST-IN-TIME ASSESMENT

To provide creativity support tailored to a given student's needs, an ICS requires a *student model* [42] that assesses the student's attributes relevant to creativity processes and outcomes throughout his or her educational activities. To prepare to conduct this research we have taken steps in this direction through related work searches that have highlighted a preliminary set of attributes that we will take into account and extend as needed. These attributes are encapsulated by Amabile's componential model of creativity and related research as follows:

Domain-Relevant Knowledge. Amabile [2] shows that the more one knows, the more opportunities there are for creativity, something the ICS student model needs to account for this in its assessment of a student's creativity. Also related to the assessment of creativity is the fact that its very definition involves the production of a novel idea or problem-solving step - the most natural way for a model to determine novelty is whether the student already possessed the knowledge related to the idea or step or if it was constructed on the spot.

Affect (Motivation and Emotion). How students feel greatly influences the creativity process and its outcomes. Thus, the ICS model will rely on the data from affective sensing devices as well as tutor variables to assess states like intrinsic motivation, central to Amabile's theory [2], as well as other affective states such as frustration (e.g., indicating Stuck!) and Flow.

Metacognition Related to Creative Thinking Styles. The third element of Amabile's theory pertains to what she terms as "creative thinking style", such as how flexible and imaginative a person is in their approach to problems, indicating the metacognitive skills required for creativity.

The ICS creativity student model will represent and infer information related to these three attributes. For the modeling of knowledge and metacognition we will build student models via established techniques (e.g., [11, 15, 28, 33]) for modeling of these attributes. Specifically, we will use cognitive and metacognitive task analysis to identify fine-grained skills needed to solve a problem (*knowledge*) and for creativity in general (*metacognition*, including, for instance, divergent thinking). These skills can be computationally represented using a *rule-based* approach that enables the system to automatically model both the target solutions and skills sets [4]. This is accomplished by tying parameters to each rule to represent the probability that the student knows the corresponding skill, which "fire" when a certain threshold is exceeded. In addition, this approach can be used to provide the backbone of a Bayesian network that makes the structure of the student knowledge and metacognitive skills explicit, as in [11]. Overall, this probabilistic approach has the advantage of recognizing that modeling student knowledge and metacognition is not a black and white process, since there is typically inherent uncertainty arising from, for instance, student slips and guesses [33] and/or lack of direct evidence on student state of interest (e.g., divergent thinking).

For the modeling of affect, initially, we will refine our existing student models developed in our work (e.g., [6]) and use their output as inputs to the ICS creativity model. These models already capture attributes that are relevant to the research at hand (e.g., interest, related to intrinsic motivation and Flow, frustration) by relying on data from the sensing devices and tutor variables. However, as mentioned above, these models do not take into account the uncertainty inherent in assessing affect as other existing affective models do (e.g., [12]) and so we will extend and/or redesign them as needed.

In order to calibrate the main ICS creativity model, as well as its knowledge, affect, and meta-cognition sub-models, we will conduct empirical studies to collect data from students (high school and college) as they interact with the target tutor while a target set of sensors captures their physiological responses. The goal behind these evaluations will be to collect a rich data set that enable us to (1) evaluate the accuracy of the student models for capturing the target student attributes and (2) analyze how student actions and student affect influence the creative process during open-ended problem solving.

Model Accuracy: To determine student model accuracy we will compare model output to a gold standard [5, 16, 29]. In the case of student knowledge, this gold standard is typically a test targeting the domain concepts. For affect and metacognition, the situation is more complicated since information on students' feelings and high-level thoughts is not readily available. Thus, we will use a two-prong approach that we have relied on in the past: (1) talk-aloud protocol by having students verbalize their thoughts and feelings (e.g., [29]) for a subset of the participants (since this is a laborious process that requires transcription and analysis of many fine-grained events), and (2) for obtaining affect information, the target system will intermittently ask students to report on their emotions (as in, [5]). Note that these techniques are only necessary during model-testing – once the model is calibrated, the self-report prompts and talk-aloud protocol are removed. To use this data to assess model accuracy, we will transcribe the talk-aloud protocols and identify metacognitive and affective events, and then use this data in conjunction with the self-report data to compare against the corresponding submodel output.

Factors Influencing Creativity: While work in psychology has provided indications of how various attributes influence creativity, the technological context of this approach affords opportunities for investigating creativity beyond traditional settings. In particular, the PI's suite of sensors provides a unique chance for extending the community's knowledge on factors that influence creativity. Thus, we will rely on the EDM techniques we have used in the past [30] and/or adopt additional ones as needed in order to mine the rich data set collected in this phase for factors influencing creativity. Specifically, relevant features will be extracted (e.g., affective states, productivity during problem solving, effort invested) and used as inputs to EDM techniques (e.g., Bayesian network parameter learning [30], logistic regression [13, 14]). This will inform how various events contribute to creativity (e.g., a student reported frustration and this was related to a low creativity time span) and the relative utility of each event to the overall creativity process. We also plan to analyze the relative utility of each sensor (as we did in [29] and [14]) in order to understand which sensors are most valuable for creativity assessment as well as what the trade offs are when not all sensors are available.

CONCLUSION

We have discussed the Intelligent Creativity Support framework and its application to the integration of ITS, affective computing, and CST to foster students' creativity. We have also outlined our research strategies for taking the next steps to implement and evaluate this approach.

REFERENCES

[[1] Aleven, V., McLaren, B., Roll, I., and Koedinger, K. (2006) Toward Meta-Cognitive Tutoring: A Model of Help-Seeking with a Cognitive Tutor. *International Journal of Artificial Intelligence in Education*, 16(2): 101- 128.

[2] Amabile, T. M. (1983) The Social Psychology of Creativity: A Componential Conceptualization. *Journal of Personality and Social Psychology*, 45(2): 357-376.

[3] Amabile, T. M., Barsade, S. G., Mueller, J. S., and Staw, B. M. (2005) Affect and Creativity at Work. *Administrative Science Quarterly*, 50(3): 367-403.

[4] Anderson, J. R. (1993) *Rules of the Mind*: Psychology Press.

[5] Arroyo, I., Cooper, D., Burleson, W., Muldner, K., and Christopherson, R. (2009) Emotion Sensors Go to School. *In Proceedings of the 14th International Conference on Artificial Intelligence in Education (AIED'09)*, Brighton, England, 17-24.

[6] Arroyo, I., Cooper, D. G., Burleson, W., Woolf, B. P., Muldner, K., and Christopherson, R. (2009) Emotion Sensors Go to School. *In Proceedings of the 2009 Conference on Artificial Intelligence in Education: Building Learning Systems that Care: From Knowledge Representation to Affective Modelling*, 17-24.

[7] Bielaczyc, K., Pirolli, P. L., and Brown, A. L. (1995) Training in Self-Explanation and Self-Regulation Strategies: Investigating the Effects of Knowledge Acquisition Activities on Problem Solving. *Cognition and Instruction*, 13(2): 221-252.

[8] Cheung, C. K., Roskams, T., and Fisher, D. (2006) Enhancement of Creativity through a One-Semester Course in University. *The Journal of Creative Behavior*, 40: 1-25.

[9] Clapham, M. M. (1997) Ideational Skills Training: A Key Element in Creativity Training Programs. *Creativity Research Journal*, 10(1): 33-44.

[10] Cohen, P. A., Kulik, J. A., and Kulik, C. L. C. (1982) Educational Outcomes of Tutoring: A Meta-Analysis of Findings. *American Educational Research Journal*, 19(2): 237-248.

[11] Conati, C., Gertner, A., and VanLehn, K. (2002) Using Bayesian Networks to Manage Uncertainly in Student Modeling. *User Modeling & User-Adapted Interaction*, 12(4): 371-417.

[12] Conati, C. and Maclaren, H. (2009) Empirically Building and Evaluating a Probabilistic Model of User Affect. *User Modeling and User-Adapted Interaction*.

[13] Cooper, D., Arroyo, I., Woolf, B., Muldner, K., Burleson, W., and Christopherson, R. (2009) Sensors Model Student Self Concept in the Classroom. *In Proceedings of the UMAP 2009, First and Seventeenth International Conference on User Modeling, Adaptation and Personalization*, Trento, Italy, 30-41.

[14] Cooper, D. G., Muldner, K., Arroyo, I., Woolf, B. P., Burleson, W., and Dolan, R. (2010) Ranking Feature Sets for Emotion Models Used in Classroom Based Intelligent Tutoring Systems. *In Proceedings of the International*

Conference on User Modeling and Adaptive Presentation (UMAP'10), 135-146.
[15] Corbett, A. T., McLaughlin, M. S., and Scarpinatto, K. C. (2000) Modeling Student Knowledge: Cognitive Tutors in High School and College. *User modeling and user-adapted interaction*, 10: 81-108.
[16] D'Mello, S. and Graesser, A. (2012) Language and Discourse Are Powerful Signals of Student Emotions During Tutoring. *Learning Technologies, IEEE Transactions on*, PP(99): 1-1.
[17] Dewett, T. and Gruys, M. L. (2007) Advancing the Case for Creativity through Graduate Business Education. *Thinking Skills and Creativity*, 2(2): 85-95.
[18] Florida, R. (2002) *The Rise of the Creative Class*. New York, NY: Basic books.
[19] Isen, A. (2004) Some Perspectives on Positive Feelings and Emotions: Positive Affect Facilitates Thinking and Problem Solving. In Antony Manstead (Eds), *The Amsterdam Symposium*. Cambridge: New York.
[20] Isen, A. M., Daubman, K. A., and Nowicki, G. P. (1987) Positive Affect Facilitates Creative Problem Solving. *Journal of Personality and Social Psychology*, 52(6): 1122-1131.
[21] Kaufman, J. C. and Beghetto, R. A. (2009) Beyond Big and Little: The Four C Model of Creativity. *Review of General Psychology*, 13(1): 1-12.
[22] Koedinger, K. R., Anderson, J. R., Hadley, W. H., and Mark, M. A. (1997) Intelligent Tutoring Goes to School in the Big City. *In Proceedings of the 7th World Conference on Artificial Intelligence in Education.*
[23] Lepper, M. R. (1988) Motivational Considerations in the Study of Instruction. *Cognition and Instruction*, 5(4): 289-309.
[24] Ma, H.(2006) A Synthetic Analysis of the Effectiveness of Single Components and Packages in Creativity Training Programs. *Creativity Research Journal*, 18(4): 435-446.
[25] Mack, R. W. (1987) Are Methods of Enhancing Creativity Being Taught in Teacher Education Programs as Perceived by Teacher Educators and Student Teachers? *The Journal of Creative Behavior*, 21(1): 22-33.
[26] Mayer, R. E. (1999) Fifty Years of Creativity Research. In R. J. Sternberg (Eds), *Handbook of Creativity*. Cambridge University Press.
[27] McCorkle, D. E., Payan, J. M., Reardon, J., and Kling, N. D. (2007) Perceptions and Reality: Creativity in the Marketing Classroom. *Journal of Marketing Education*, 29(3): 254-261.
[28] Mitrovic, A. (2012) Fifteen Years of Constraint-Based Tutors: What We Have Achieved and Where We Are Going. *User Modeling and User-Adapted Interaction*, 22(1-2): 39-72.
[29] Muldner, K., Burleson, W., and VanLehn, K. (2010) “Yes!”: Using Tutor and Sensor Data to Predict Moments of Delight During Instructional Activities. *In Proceedings of the User Modeling, Adaptation, and Personalization (UMAP'10)*, 159-170.
[30] Muldner, K., Burleson, W., Van de Sande, B., and VanLehn, K. (2011) An Analysis of Students’ Gaming Behaviors in an Intelligent Tutoring System: Predictors and Impacts, User Modeling and User Adapted Interaction. *Journal of Personalization Research, Special Issue on Data Mining for Personalized Educational Systems*: 99-135.
[31] NGA and CCSSO. (2012) *Common Core State Standards Initiative: Preparing America's Students for College & Career*. Available from: www.corestandards.org.
[32] Picard, R. (1997) *Affective Computing*. Cambridge, MA: MIT Press.
[33] Reye, J. (2004) Student Modeling Based on Belief Networks. *International Journal of Artificial Intelligence in Education (IJAIED)*, 14: 1-33.
[34] Robbins, T. L. and Kegley, K. (2010) Playing with Thinkertoys to Build Creative Abilities through Online Instruction. *Thinking Skills and Creativity*, 5(1): 40-48.
[35] Runco, M. A. (2004) Creativity. *Annual Review of Psychology*, 55: 657-687.
[36] Scott, G., Leritz, L. E., and Mumford, M. D. (2004) The Effectiveness of Creativity Training: A Quantitative Review. *Creativity Research Journal*, 16(4): 361-388.
[37] Self, J. (1998) The Defining Characteristics of Intelligent Tutoring Systems Research: Itss Care, Precisely. *International Journal of Artificial Intelligence in Education*, 10: 350-364.
[38] Shute, V. and Psotka, J. (1996) Intelligent Tutoring Systems: Past, Present, and Future. In (Eds), *Handbook of Research for Educational Communications and Technology*. Macmillan: New York.
[39] Trilling, B. and Fadel, C. (2009) *21st Century Skills: Learning for Life in Our Times*. SF: Jossey-Bass.
[40] Tripathi, P. and Burleson, W. (2012) Predicting Creativity in the Wile: Experience Sample and Sociometric Modeling of Teams. *In Proceedings of the Conference on Computer Supported Cooperative Work (CSCW'12)*.
[41] Vance, E. (2007) Colleges Should Teach Broader Skills to Prepare Students for Work Force, Report Says. *The Chronicle of higher education*.
[42] VanLehn, K. (1988) Student Modeling. In M. Polson and J. Richardson (Eds), *Foundations of Intelligent Tutoring Systems* Erlbaum: Hillsdale, NJ. 55-78.
[43] VanLehn, K., Lynch, C., Schulze, K., Shapiro, J. A., Shelby, R., Taylor, L., Treacy, D., Weinstein, A., and Wintersgill, M. (2005) The Andes Physics Tutoring System: Five Years of Evaluations. *In Proceedings of the Artificial Intelligence in Education Conference (AIED'05)*, 678-685.
[44] Vygotsky, L. S. (2004) Imagination and Creativity in Childhood. *Journal of Russian & East European Psychology*, 42(1): 7-97.
[45] Wince-Smith, D. L., *The Creativity Imperative: A National Perspective*. 2006. 12 - 14.

Design and Evaluation for *Safe* Improvisation in EMS Technologies

Steven R. Haynes, David R. Mudgett, John M. Carroll
College of Information Sciences & Technology
The Pennsylvania State University
University Park, PA 16803
{shaynes, dmudgett, jcarroll} @ist.psu.edu

ABSTRACT

In this paper we explore the concept of *safe improvisation* and propose principles for design and evaluation of interactive technologies to support improvisational action in safety-critical domains. A particular focus here is on the interactive technologies used in emergency medical services (EMS). In EMS improvisation is necessitated by events, but discouraged within its culture. Designing and evaluating for this conflict is central to development of technologies that are at once institutionally acceptable, usable in the field, and safe to employ.

Author Keywords

Improvisation, evaluation, design, emergency medical services.

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms

Human Factors.

INTRODUCTION

In this paper we explore how technologies can be designed and then assessed in terms of their ability to support *safe* improvisation. Many safety-critical activities such as piloting an aircraft, driving a car, managing a power plant, administering radiopharmaceuticals, combat operations, and emergency response happen within an organizational and social context that places strict limits on the scope of human behavior. At the same time, the flow of events in particular circumstances can necessitate improvisational action, actions that violate protocols of standard practice but are required to respond effectively to emergent situations. The protocols, operational guidelines, and policies that bound activity in these domains are there for a good reason. They are designed to ensure the safety of both the performer and those directly and indirectly affected by their actions.

We consider improvisation as a sub-type of creativity. Improvisation is creative action situated in an activity being practiced, in real-time, and within a particular and often unique context. Improvisation is emergent, necessitated and shaped by events rather than planned. This differentiates improvisation from creativity in the arts, for example, where creative action is (or can be) both more deliberate and more 'organic'. The relationship between necessity and improvisation is particularly salient in safety-critical domains, where task performers typically intend and prefer to operate within established, regulated boundaries, and where deviation from protocol is a sign that something has somehow gone wrong.

Our focus here is on the role of interactive technologies in situations requiring potentially unsafe improvisations, and in particular how these technologies can be designed and then evaluated in terms of their ability to support, guide, and enhance the quality of improvisation while ensuring the safety of those involved. Our study domain is emergency medical services, in particular the technology used in EMS, and how improvisation is both necessitated and governed by EMS practice. Our analysis is based both on literature and on experience; one author is a volunteer emergency medical technician (EMT) within a regional EMS service. Note that our experience and our analysis focus on EMS practices, technologies, and improvisation in the United States. We acknowledge that this context and EMS practice may differ widely from nation to nation and from culture to culture.

EMERGENCY MEDICAL SERVICES

Emergency medical services (EMS) are the forward-facing extreme of the patient healthcare continuum. Among the most pressing challenges to effective EMS are information gathering, information capture and integration, and information use. Emergency medical services generally perform three types of service: responding to emergencies (in the United States they are typically initiated by a telephone call to 9-1-1); routine and critical-care transports between a patient's residence and a healthcare facility, or between healthcare facilities; and ambulance standbys at both planned (e.g. a state fair) and unplanned (e.g. a

CHI'13, April 27 – May 2, 2013, Paris, France.
ACM 978-1-XXXX-XXXX-X/XX/XX

structure fire) events. Here the focus is on 9-1-1 emergencies and the intensive information capture, creation, communication, use, and management activities that attend these events.

Ambulance emergency responses (9-1-1 calls) are generally categorized in one of two ways: *medical*, such as a diabetic emergency or myocardial infarction (heart attack); or *trauma*, such as a broken leg at a sporting event, or multi-system injury at a motor vehicle accident. In either case a very large quantity of information is required to be captured in a relatively small amount of time. This includes dispatch information, the location and nature of the call; patient information both demographics and healthcare-relevant; treatment information, what was done to address the patient's condition; and transport information including emergency status (lights and sirens), transport times, and mileage.

Paramedics and EMTs use a variety of interactive technologies to support emergency response, situational and patient assessment, and patient care. Information gathering practices are however, still largely based on verbal communication with patients themselves, with their family members, and with bystanders who might have witnessed the event. Emergency information capture and integration is still largely done with pen and paper. Because of this, we view EMS as an information-intensive domain that is so-far under-served by available information technologies.

Figure 1 - A Patient Monitor Used in EMS

Interactive technologies in EMS must be small and portable, durable, easy to learn and use, and secure. Information is central to effective response and patient care and is used intensively in communication, navigation, situational awareness, clinical decision making for patient assessment and treatment, and patient care reporting. Many EMS services, especially smaller regional and rural services, are relatively under-funded, making information technology investment a significant challenge.

A 2007 report from the US National Academy of Sciences [1] identified a number of pressing challenges to achieving more effective pre-hospital emergency care. These include especially that EMS is a relatively under-resourced layer in the continuity of care that begins in the field and extends through emergency departments to more specialized care facilities. The report also highlights the importance of information and communications systems integration as a key enabler of more seamless and cost-effective information sharing across different levels and functions in the health care system.

Current and emerging technologies considered especially relevant to EMS include increasingly capable mobile devices, voice control interfaces, location-aware services, and cloud computing. On the horizon are technologies such as mobile ultrasound, and non-invasive blood testing, both of which represent powerful tools for enhanced patient assessment in the field. The potential for integration of these technologies is especially promising. Real-time communication of patient imaging and blood test results between healthcare providers in the field and hospital emergency departments could be a life-saving advantage in many critical care cases.

Still needed though is a reference design model of an integrated EMS architecture focused on delivering critical and time-sensitive information to pre-hospital care providers. Components of the architecture should include a service layer supporting common EMS capabilities, and a presentation layer making use of emerging mobile technologies and alternative interfacing styles. Perhaps most important is to develop ways that the potential for such technologies can be understood and measured in the context of actual EMS operations.

The role of improvisation in EMS is a difficult and conflicted topic. On one hand, performers in the United States are bound by a detailed set of protocols (see for example: [2]) that determine the scope of practice (what they can do) as well as how they are allowed to do it. On the other, the culture of EMS, at least in principle, emphasizes "the thinking EMT" who should be capable of responding effectively to emergent problems arising in the sometimes chaotic and always time-constrained course of an ambulance call. Improvisation in EMS is driven by necessity and sometimes lives depend on effective improvisation. Improvisations, however, can also expose healthcare providers, patients, and others to a wide range of risks.

DESIGN & EVALUATION FOR SAFE IMPROVISATION

In this section we identify a set of nine principles for designing and evaluating interactive technologies in terms of whether and how well they support safe improvisation. We view principles such as these as useful resources for both *formative* evaluation in the design phases and *summative* evaluation once a technology is built and deployed. These principles are derived from both the literature of improvisation and from experience in the field of EMS. The proposed principles are intended to provoke discussion. Ongoing and future work will involve expanding, refining, and validating this list through studies in the field with a local regional EMS organization.

Nine Principles for Safe Improvisation

Because EMS and other safety-critical domains are often protocol-driven, with established procedures for task conduct [3], interactive tools should explicitly implement **activity themes, genres, rules, standards** into task support. Such task support technology would have the ability to monitor performer actions in terms of its deviation from accepted norms and warn when a deviation might result in an unsafe improvisation.

A common theme in studies of creativity and improvisation is that they are a form of expert performance requiring deep **knowledge of tools, materials, processes** [4]. This knowledge is a function of training and practice, but might also be provided or cued in real-time to a performer of a situated activity. Notification-based 'help' systems such as these are notoriously challenging to implement so that fluid task support is provided without needless interruptions.

Incorporating **risk awareness and risk analysis** into improvisational activity support is essential to ensuring that tools providing such support also ensure the safety of the performer. This might involve, for example, a notification facility that warns a performer engaged in a task where deviations from accepted practice carry known risks.

Some researchers have highlighted the role of **organizational memory** in supporting improvisation [5]. Improvisations in EMS, for example, occur within organizations that possess deep knowledge of local environmental factors and how they both constrain and enable safe improvisations. Such organizational memory might be made available to performers informally, as an online forum, for example, or more formally as part of an organization's explicit standard operating procedures.

A particularly challenging criterion is the ability of interactive systems to promote **divergent thinking** in the course of task performance. Divergent thinking is at the core of improvisation [6]. The challenge to technologies in safety-critical domains is to support divergent thinking while concurrently monitoring the performers drift from established practice and into modes of behavior that might endanger the stakeholders involved in an EMS scenario.

Though we include it here as a proposal, encouraging real and perceived **freedom to act** within safe boundaries may not be within the scope of an interactive technology's design. Freedom to improvise is likely more a function of organizational, institutional, and professional culture, which in EMS must account for the direct risks of improvisation to the patient and to oneself, and indirect risks arising from the regulatory and legal contexts.

An advanced, perhaps idealized future capability for improvisation support technologies is helping to identify or even predict situations where **necessity** will dictate an improvised response. Time is often critical in emergency medicine and real value might be gained from knowing in advance when an established practice is not going to result in a positive outcome.

As mentioned earlier one aspect of improvisation that distinguishes it from other forms of creativity is that composition and performance (to use terms from music) occur simultaneously or near-simultaneously [5]. Supporting simultaneous **composition and action** suggests both an information filtering and system performance criterion. Actors must have the ability to quickly obtain and integrate information related to the current situation then formulate this information as a safe plan of action.

A final proposed criterion for technologies supporting safe improvisation is their ability to support **situational decision-making.** In a sense, of the criteria that have been discussed are in service of operational and clinical decision-making tasks.

A Role for Design Rationale?

One emergent idea from our analysis to-date relates to the potential role of *design rationale* in helping to implement these principles. Design rationale is the reasoning behind a design. This can be and has been thought of in a variety of ways. Design rationale can be a description of the space of issues, possibilities, and decision criteria that framed a design process [7]. It can be a structured account of the *particular* issues, possibilities, and decision criteria that comprised a given design process [8]. Design rationale can be a tradeoff analysis of the features of a design with respect to given criteria [9], or a tradeoff analysis the *usage contexts* for an envisioned design as in scenario-based design [10].

These activities and the products that result provide a rich knowledge base that helps relate the form of a particular design to the reasoning process that led to its realization. Designers of technologies for use in safety-critical domains must be ever-mindful of how their designs ensure safe use of the devices they create. How they evaluate safety within a design space as a whole, and within the context of a particular design decision or design 'move' describes the safe boundary of design use. This knowledge is brought to the surface in the design process, but then once a decision is made it is typically made invisible, despite being inscribed into the technology. Making this knowledge more accessible to users of a technology may help them see how the safety parameter was included in activity support.

In Figure 2 below we show some theorized relationships between these evaluation criteria. We suggest that some of the enablers and determinants of improvisational practice supervene, or are otherwise determined by the existence and scale of other related factors. Again, understanding the nature of both these factors and their relations will be explored in future empirical work.

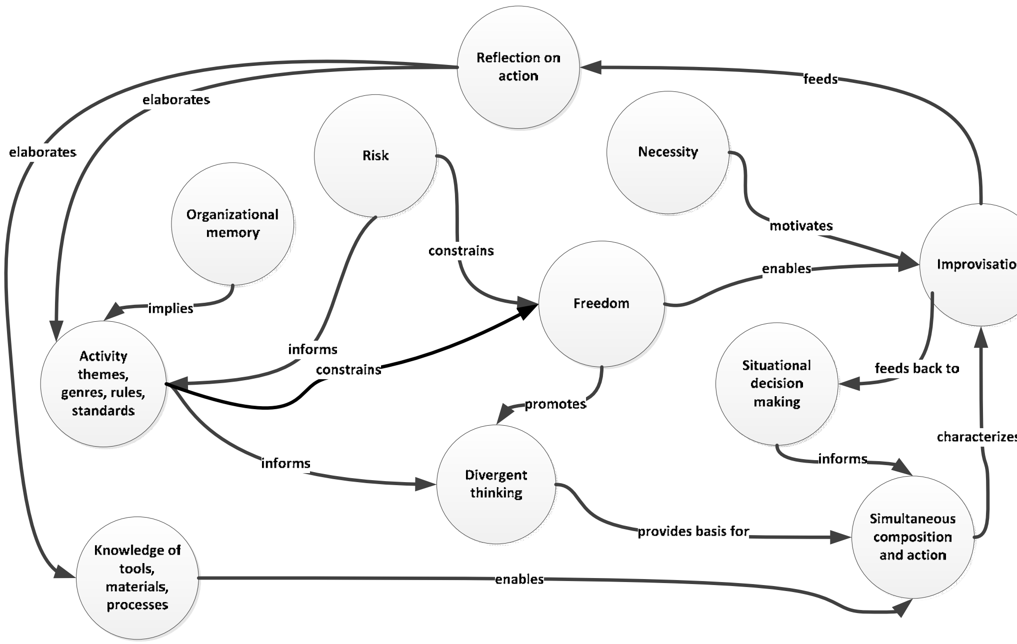

Figure 2 - Evaluation criteria for interactive technology supporting safe improvisation.

MOVING FORWARD

As information systems become increasingly pervasive our approach to design must adapt to include new principles and constraints introduced by new domains. Future generations of technology users will demand that the technologies they use acknowledge and inscribe the cultural norms that govern their work. In domains where risks are significant, failures consequential, and improvisation necessary, information systems should act as a kind of 'third hand' to help guide and govern the boundaries of human action. Technologies in safety-critical domains present special requirements for improvisational use.

We are working to better understand these requirements and how they can be applied when creating new technologies. Much of the work so far has been analytic. Moving forward we are planning a series of empirical studies that we hope will extend the range of our understanding and help ground design prescriptions in the practices for which they are intended. Our current focus is on a qualitative field study examining the role of improvisation in the work of EMTs and paramedics in a regional emergency medical service.

REFERENCES

[1] IOM-US, *Emergency medical services at the crossroads*. Washington, D.C.: National Academies Press, 2007.

[2] PA-DOH, "Pennsylvania Statewide Basic Life Support Protocols," ed. Harrisburg, PA: Pennsylvania Department of Health, 2011.

[3] S. R. Haynes, W. A. Schafer, and J. M. Carroll, "Leveraging and Limiting Practical Drift in Emergency Response Planning," in *40th Annual Hawaii International Conference on System Sciences*, Waikoloa, HI, USA, 2007, pp. 200-208.

[4] D. Schön and J. Bennett, "Reflective Conversation with Materials," in *Bringing Design to Software*, T. Winograd, Ed., ed: Addison-Wesley, 1996, pp. 171 - 189.

[5] C. Moorman and A. S. Miner, "Organizational Improvisation and Organizational Memory," *Academy of Management Review,* vol. 23, pp. 698-723, 1998.

[6] C. Dym, A. Agogino, O. Eris, D. Frey, and L. Leifer, "Engineering Design Thinking, Teaching, and Learning," *Journal of Engineering Education,* vol. 94, pp. 103-114, 2005.

[7] A. MacLean, R. M. Young, V. M. E. Bellotti, and T. Moran, "Questions, Options, and Criteria: Elements of Design Space Analysis," in *Design Rationale: Concepts, Techniques and Use*, T. P. Moran and J. M. Carroll, Eds., ed Mahwah, NJ: Lawrence Erlbaum, 1996, pp. 21-51.

[8] Kunz, W. and H. W. J. Rittel (1970). Issues as elements of information systems (Working Paper 131), Institute for Urban and Regional Development, University of California, Berkeley.

[9] J. Lee and K.-Y. Lai, "What's in Design Rationale?," in *Design Rationale: Concepts, Techniques and Use*, T. P. Moran and J. M. Carroll, Eds., ed Mahwah, NJ: Lawrence Erlbaum, 1996, pp. 21-51.

[10] J. M. Carroll and M. B. Rosson, "Getting around the task-artifact cycle: How to make claims and design by scenario," *ACM Transactions on Information Systems,* vol. 10, pp. 181-212, 1992.

Creativity Support in Authoring and Backtracking

Brad A. Myers, Stephen Oney,
Human-Computer Interaction Inst.
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213
{bam, soney}@cs.cmu.edu

YoungSeok Yoon,
Inst. for Software Research
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213
youngseok@cs.cmu.edu

Joel Brandt
Adobe Research
San Francisco, CA 94103
joel.brandt@adobe.com

ABSTRACT

The "Natural Programming" group has been working for 15 years on making it easier for all kinds of programmers to be creative when writing software. Recently, one focus has been enabling "end-user programmers" (EUPs) such as interaction designers to more easily author interactive behaviors for the web. In a separate project, we are adding features to a code editor to support "backtracking" — undoing operations to partially or fully restore the code to a previous state — since creative exploration usually involves both moving forward to new designs and going backwards to retract some or all of the design that is not desired. In all of these projects, we seek to measure both the usability of our tools, and their effectiveness at fostering creativity.

Author Keywords

Interactive Behaviors; End-User Programming; Exploratory Programming; Backtracking; Natural Programming; Creativity.

ACM Classification Keywords

H.5.2 [Information Interfaces and Presentation]: User Interfaces – Interaction styles; H.1.2 [User/Machine Systems]: Software psychology; D.2.6 [Programming Environments]: Integrated environments

General Terms

Human Factors; Design; Measurement.

INTRODUCTION

Buxton quotes Linus Pauling as having said: "The best way to have a good idea is to have lots of ideas" [4, p. 121]. Design education places much emphasis on this strategy [6] and research suggests that exploring multiple ideas helps improve creativity [2, 8]. Moreover, creativity theory suggests the need to produce a plethora of ideas in order to arrive at the creative ones, a concept called *ideational fluency* [13]. Donald Schön, one of the most influential design theorists [11], characterizes the creative process as a conversation with materials [32]. In this conversation, designers advance the work by reflecting both in and on their actions, and by engaging with materials that specifically support conceiving and refining ideas as well as with the target material a product will be made out of. Designers reflect *in* action by evaluating and experimenting with what they are working on while they are working on it. A similar observation was made by Rosson & Carroll when studying Smalltalk programmers: a key way in which programmers create programs is to write some code, see if it does what is desired, and if not, entirely or selectively remove part of what they have created; a process they call "debugging into existence" [31]. When the process of trying out designs is embodied in writing software, it has been called *exploratory programming* [33] or *opportunistic programming* [3]. Such explorations require both forward and backward moves: forward to create new software, and backwards so the code returns at least partially to the way it was previously, either by removing inserted code or by restoring removed code. We call such backward moves *backtracking*. Besides directly helping programmers remove unwanted edits, we claim that programmers will feel more comfortable exploring if they know they have effective tools for backtracking.

The Natural Programming Project [24] tries to make programming easier by making it more *natural*, by which we mean closer to the way people think about their tasks. One way to define programming is the process of transforming a mental plan into one that is compatible with the computer [14]. The closer the programming language is to the developer's original plan, the easier this refinement process will be [12]. We have adapted a variety of HCI techniques to help understand and evaluate how developers program and use novel and conventional development tools, including Contextual Inquiry field studies [19], surveys [25], heuristic analysis and cognitive walkthrough [10], lab usability studies [9], paper prototyping [20], and *A vs. B* user studies [17].

Over the years, a number of our tools have particularly focused on the issue of *creativity*, especially as it relates to professional programmers and also "end-user programmers" (EUPs) [15], who are people who program in order to achieve some goal other than the programming itself, for example, interaction designers testing out an idea. Programmers and interaction designers often need to be creative, and it would be useful to have a way to measure the extent to which their development tools enable creativity.

To appear in:
ECSE 2013: evaluation methods for
creativity support environments
Workshop at CHI'13, April 28, 2013, Paris, France.

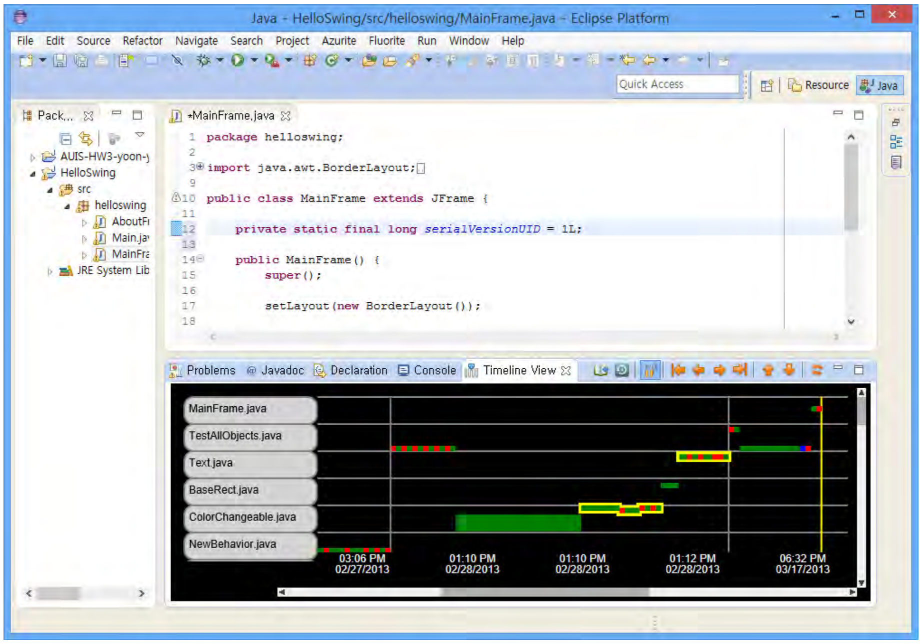

Figure 1: The current interface of AZURITE for the Eclipse IDE. The top window is the Eclipse code view, and the bottom timeline visualization shows insert (green), delete (red) and replace (blue) operations for each of the files. A vertical gray line divides the two consecutive editing sessions. The yellow vertical line at the right shows the current time. The user can select certain operations (here shown with a yellow outline), either by clicking on them, or querying in the code window for all the operations affecting a range of code, and then can selectively undo only those operations.

Current measures of creativity that we have used are all *indirect* – measuring other things that may be said to correlate with creative behavior. For example, in an early paper, we measured the number of different designs that users were able to create [18], and recently we have proposed that measuring how quickly users can move from one design to another would correlate with the success of the creativity support. However, it would certainly be useful to have a more *direct* measure of the level of creativity supported by our tools. For example, we might compare our AZURITE tool (see below) with a "preview" system in graphical editing tasks [34] or examine the tools with "sketch" like interactivity [35].

OVERVIEW OF OUR CREATIVITY SUPPORT TOOLS

Authoring

We have worked on many different kinds of authoring tools that can be considered "creativity support environments" (CSEs), at least for professional programmers and EUPs. For example, we developed many interactive tools to enable user interfaces to be created with little or no programming by user interface specialists. Some examples are Peridot [22] for creating controls (widgets like menus and scroll bars), Gamut [21] for defining behaviors by example, Silk [18] for sketching interfaces and having them automatically converted into code, HANDS [29] which is a novel programming language for kids, and Citrus [16] which is a toolkit for creating user interfaces for structured data.

Our current project for authoring is called EUCLASE, and is based on research on how interaction designers naturally express user interface behaviors, such as how the objects on the screen respond to the user [30]. We also studied how designers collaborate and express their ideas [28]. We then used these results to create a new JavaScript toolkit for creating interactive behaviors for the web, called ConstraintJS, which supports constraints combined with state diagrams [27]. Now, we are working on an interactive tool which combines a spreadsheet-like user interface with state diagrams, which we feel will enable interaction designers to more easily (and creatively) be able to author interactive behaviors themselves [26].

Backtracking

There is a large body of work and many research and commercial tools directed at making it easier for people to move forward from their ideas to designs to implementations, but there is surprisingly little support for directly helping people explore multiple variations (besides sketching on paper [4]). In particular, very little is available to help today's developers *backtrack*, even though developers report that backtracking is often required [36]. For example, modern integrated development environments (IDEs) do not utilize any of the sophisticated undo mechanisms that have been investigated through the years (e.g., [1, 5, 7, 23]), and only provide a simple linear undo model. As a result, developers cannot easily undo the changes that they made some time ago, or changes that are interleaved with edits that are still desired, but only can undo the most recent changes in the command history. Also, when the developer undoes several steps backwards and makes a new change from that point, all the previously undone commands are discarded and cannot be redone, because the undo model only keeps a linear list instead of a command history tree. In our previous survey [36], programmers reported that they use undo mostly to remove typos or repair minor mistakes in the very last edits made. Another possible way to backtrack is to use a Version Control System (VCS), but this is not always adequate: it only works if the user thought to commit the desired version, which may not always be the case, and it is often too heavy-weight for small experiments. Furthermore, if the user has made edits that need to be retained mixed in with the edits to be backtracked, neither linear undo nor version control can be used.

We are developing a plug-in called AZURITE for the Eclipse IDE, which enables users to more easily backtrack (see Figure 1). It allows users to perform *selective undo* of only the desired operations, so users can choose exactly which operations should be undone. We address two main problems of providing selective undo in the context of text editing of code: first, we provide a way to deal with *conflicts* when a later edit overlaps an earlier one, and to effectively ask the user's intent when the conflicts cannot be resolved automatically. Second, we provide a novel way for users to *find* the operations they want to undo using a timeline visualization of the code editing history (see Figure 1). In addition, unlike other existing undo models, our selective undo model allows users to select multiple edit operations at the same time, which is very effective at minimizing the occurrence of unresolvable conflicts.

Although the timeline visualization provides a way to select multiple operations, it becomes difficult to manually select the right operations to undo as the history gets bigger. Therefore, we are working on a sophisticated *search* mechanism to allow users to easily find the operations that they want to undo using whatever they remember about those operations. We already support what we found to be the most desired operation [36]: users can search for all edits performed on a particular area of code, and undo them. We also allow users to find code which *used* to exist but has subsequently been deleted. History search not only helps the users to select the right operations to undo, but also minimizes the irresolvable conflicts because the conceptually-related edits are likely to be performed on the same area of code, and thus they usually have conflicts only among themselves. In the future, we propose to also support other interesting ways to search the history, including to find the times when a particular line, method, or class was edited; to find the last time the application compiled without error or was run without raising an exception; to find when a particular editing operation was performed (e.g., "a refactor using Extract Superclass"); etc.

Although currently implemented in the context of a code editor, AZURITE should be directly applicable for any text editor, such as Word or Pages. We feel that our backtracking ideas would also transfer to other kinds of editors as well, including design programs like Photoshop, Illustrator or even PowerPoint and Keynote, where users now need to backtrack but have little support.

FUTURE EVALUATIONS

We have been building and evaluating our research prototypes iteratively, with evaluation results driving the design of future prototypes. Currently, the main focus of the evaluations is whether the tools are *usable* by the intended audiences, and whether users are *effective* at performing the tasks we are attempting to support. These evaluations use conventional usability think-aloud and *A vs. B* lab studies. In the future, we plan to deploy our prototypes to understand how they are used in practice.

We would also like to measure the extent to which our prototypes help users be more creative. With EUCLASE, we can measure whether it helps designers create novel interactive behaviors. With AZURITE, which records everything that users do, we can have a direct measure of how often people make alternative designs, and how often they backtrack. However, both of these are indirect measures, so we are interested in using other evaluation techniques and measures as well.

ACKNOWLEDGMENTS

Funding for this research comes in part from Adobe, in part from the Korea Foundation for Advanced Studies (KFAS), and in part from NSF grant IIS-1116724. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of Adobe, KFAS or the NSF.

REFERENCES

1. Berlage, T., "A Selective Undo Mechanism for Graphical User Interfaces Based on Command Objects." *ACM Transactions on Computer Human Interaction*, 1994. **1**(3): pp. 269-294.
2. Brandt, J., *et al.*, "Example-Centric Programming: Integrating Web Search into the Development Environment," in *CHI'2010: ACM Conference on Human Factors in Computing Systems*, April, 2010. Atlanta, GA. pp. 513-522.
3. Brandt, J., *et al.*, "Opportunistic Programming: Writing Code to Prototype, Ideate, and Discover." *IEEE Software*, 2009. pp. 18-24.
4. Buxton, B., *Sketching User Experiences: Getting the Design Right and the Right Design.* 2007, San Francisco, CA: Morgan Kaufmann.
5. Cass, A.G. and Fern, C.S.T., "Modeling Dependencies for Cascading Selective Undo.," in *IFIP INTERACT 2005 Workshop on Integrating Software Engineering and Usability Engineering*, 2005.
6. Cross, N., *Design Thinking: Understanding How Designers Think and Work [Paperback].* 2011, Berg Publishers.
7. Cubitt, T., *Undo-Tree.El Version 0.2 for Emacs.* 2010. http://www.dr-qubit.org/undo-tree/undo-tree-0.2.el.
8. Dow, S.P., *et al.*, "Parallel Prototyping Leads to Better Design Results, More Divergence, and Increased Self-

Efficacy." *ACM Trans. Comput.-Hum. Interact.*, 2010. **17**(4): pp. 1-24.
9. Ellis, B., Stylos, J., and Myers, B. "The Factory Pattern in Api Design: A Usability Evaluation," in *International Conference on Software Engineering (ICSE'2007).* 2007. Minneapolis, MN: pp. 302-312.
10. Faulring, A.*, et al.*, "Agent-Assisted Task Management That Reduces Email Overload," in *International Conference on Intelligent User Interfaces: IUI'2010*, February, 2010. Hong Kong, China. pp. 61–70.
11. Forlizzi, J., Zimmerman, J., and Stolterman, E. "From Design Research to Theory: Evidence of a Maturing Field," in *International Assoc. of Societies of Design Research Conference.* 2009. Seoul, Korea: IASDR.
12. Green, T.R.G., "Cognitive Dimensions of Notations," in *People and Computers V*, A. Sutcliffe and L. Macaulay, Editors. 1989, Cambridge University Press. Cambridge.
13. Guilford, J.P., *Intelligence, Creativity and Their Educational Implications.* 1968, Robert R. Knapp.
14. Hoc, J.-M. and Nguyen-Xuan, A., "Language Semantics, Mental Models and Analogy," in *Psychology of Programming*, J.-M. Hoc*, et al.*, Editors. 1990, Academic Press. London. pp. 139-156.
15. Ko, A.J.*, et al.*, "The State of the Art in End-User Software Engineering." *ACM Computing Surveys*, 2011. **43**(3): pp. Article 21, 44 pages.
16. Ko, A.J. and Myers, B.A. "Citrus: A Toolkit for Simplifying the Creation of Structured Editors for Code and Data," in *UIST'05: ACM Symposium on User Interface Software and Technology.* 2005. Seattle, WA: pp. 3-12.
17. Ko, A.J. and Myers, B.A., "Finding Causes of Program Output with the Java Whyline," in *CHI'2009: Human Factors in Computing Systems*, April 4-9, 2009. Boston, MA. pp. 1569-1578.
18. Landay, J.A., *Interactive Sketching for the Early Stages of User Interface Design.* Ph.D. Thesis, Computer Science Department, Carnegie Mellon University, 1996, CMU-HCII-96-105.
19. LaToza, T.D. and Myers, B., "Developers Ask Reachability Questions," in *ICSE'2010: Proceedings of the International Conf. of Software Engineering*, May 2-8, 2010. Capetown, South Africa. pp. 185-194.
20. LaToza, T.D. and Myers, B.A., "Visualizing Call Graphs," in *VL/HCC'2011: IEEE Symposium on Visual Languages and Human-Centric Computing*, Sept. 18-22, 2011. Pittsburgh, PA. pp. 117-124.
21. McDaniel, R.G. and Myers, B.A. "Getting More out of Programming-by-Demonstration," in *Proceedings CHI'99: Human Factors in Computing Systems.* 1999. Pittsburgh, PA: pp. 442-449.
22. Myers, B.A. "Creating Dynamic Interaction Techniques by Demonstration," in *CHI+GI'87: Human Factors in Computing Systems.* 1987. Toronto, Ont., Canada: pp. 271-278.
23. Myers, B.A. and Kosbie, D., "Reusable Hierarchical Command Objects," in *CHI'96: Human Factors in Computing Systems*, April 14-18, 1996. Vancouver, BC, Canada. pp. 260-267.
24. Myers, B.A., Pane, J.F., and Ko, A., "Natural Programming Languages and Environments." *Communications of the ACM*, 2004. **47**(9): pp. 47-52.
25. Myers, B.A.*, et al.*, "How Designers Design and Program Interactive Behaviors," in *2008 IEEE Symposium on Visual Languages and Human-Centric Computing, VL/HCC'08*, Sept 15-18, 2008. Herrsching am Ammersee, Germany. pp. 185-188.
26. Oney, S., Myers, B., and Zimmerman, J. "Visions for Euclase: Ideas for Supporting Creativity through Better Prototyping of Behaviors," in *ACM CHI 2009 Workshop on Computational Creativity Support.* 2009. Boston, MA:
27. Oney, S., Myers, B.A., and Brandt, J., "Constraintjs: Programming Interactive Behaviors for the Web by Integrating Constraints and States," in *UIST'2012: ACM Symposium on User Interface Software and Technology*, October 7-10, 2012. Cambridge, MA. pp. 229-238.
28. Ozenc, K.*, et al.*, "How to Support Designers in Getting Hold of the Immaterial Material of Software," in *CHI'2010: Human Factors in Computing Systems*, April 10-15, 2010. Atlanta, GA. pp. 2513-2522.
29. Pane, J.F., Myers, B.A., and Miller, L.B., "Using Hci Techniques to Design a More Usable Programming System," in *IEEE 2002 Symposia on Human Centric Computing Languages and Environments (HCC 2002),* , September 3-6, 2002. Arlington, VA. pp. 198-206.
30. Park, S., Myers, B., and Ko., A. "Designers' Natural Descriptions of Interactive Behaviors," in *2008 IEEE Symposium on Visual Languages and Human-Centric Computing, VL/HCC'08.* 2008. Herrsching am Ammersee, Germany: pp. 185-188.
31. Rosson, M.B. and Carroll, J.M., "The Reuse of Uses in Smalltalk Programming." *ACM Transactions on Computer-Human Interaction*, 1996. **3**(3): pp. 219-253.
32. Schön, D., *The Reflective Practitioner.* 1983, London: Temple Smith.
33. Sheil, B., "Environments for Exploratory Programming." *Datamation*, 1983. reprinted in in "Papers on Interlisp-D," Sheil, B.A. and Masinter, L.M., eds., Xerox PARC Tech Report CIS-5.
34. Terry, M. and Mynatt, E.D., "Side Views: Persistent, on-Demand Previews for Open-Ended Tasks," in *UIST'2002: 15th annual ACM symposium on User interface software and technology*, 2002. ACM: Paris, France. pp. 71-80.
35. Yamamoto, Y. and Nakakoji, K., "Interaction Design of Tools for Fostering Creativity in the Early Stages of Information Design." *International Journal of Human-Computer Studies*, 2005. **63**(4-5): pp. 513-535.
36. Yoon, Y. and Myers, B.A., "An Exploratory Study of Backtracking Strategies Used by Developers," in *Cooperative and Human Aspects of Software Engineering (CHASE'2012), An ICSE 2012 Workshop.*, June 2, 2012. Zurich, Switzerland. pp. 138-144.

A Tiered Evaluation Framework for Reality-Based Creativity Support Environments

Orit Shaer, Consuelo Valdes
Wellesley College
106 Central St, Wellesley, MA
{oshaer, cvaldes} @wellesley.edu

ABSTRACT

Reality-based interfaces (RBIs) such as tabletop and tangible user interfaces offer unique opportunities for supporting creativity. In this paper, we present a multi-tiered evaluation framework for reality-based creativity support environments and describe its application in the context of college-level science education. The proposed framework consists of three layers that examine the usability, usefulness, and impact of an environment on creative problem solving processes. Drawing upon the existing body of work in the area, our framework documents a mixed-method approach and provides guidance for the evaluation of reality-based creativity support environments.

Author Keywords

Reality-based interaction; creativity support; tabletop interaction; collaborative learning.

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms

Human Factors; Design; Measurement.

INTRODUCTION

Over the past two decades, HCI research has generated a broad range of interaction styles that move beyond the desktop into new physical and social contexts. Key areas of innovation have been tabletop, tangible, and embodied user interfaces. These interaction styles share an important commonality: leveraging users' existing knowledge and skills of interaction with the real non-digital world, thus they can be unified under the umbrella of Reality-Based Interfaces (RBIs) [4]. Building upon ideas from embodied cognition, RBIs offer a more natural, intuitive, and accessible form of interaction that reduces the mental effort required to learn and operate a computational system [4].

Given the potential of RBIs, numerous research prototypes have explored how these emerging interaction styles will impact education. Several studies have examined the effects of RBIs on learning, investigating the benefits and deficits of reality-based interaction in the contexts of formal and informal learning. However, most of these studies have focused on children. To date, little research has been devoted to investigating the strengths and limitations of utilizing RBIs for promoting creative thinking in scientific problem solving at the high school or college levels.

Our focus is on investigating the application of reality-based interaction for promoting creative thinking in college-level science education. Creative thinking in the context of science education is a multicomponent process, which is influenced by group interactions and social context [1]. The creative process that leads to an individual insight (i.e. unit of discovery [9]) includes at least three diverse, but testable elements [1]: 1) divergent thinking – the ability to generate and accept many ideas related to a problem; 2) convergent thinking – the ability to focus and mentally evaluate ideas; and 3) analogical thinking – the ability to understand a novel idea in terms of existing knowledge. In our work, we have explored the application of reality-based interaction for enhancing creative thinking in data-intensive areas such as college-level genomics [11, 12, 13] and phenology [15].

In this position paper we describe a multi-tiered evaluation framework for understanding the strengths and limitations of reality-based creativity support environments [13]. We applied this framework in the evaluation of three reality-based interfaces for inquiry-based college level science learning: G-nome Surfer [11, 12, 13] – a tabletop user interface for collaborative exploration of genomic information; GreenTouch [15] – a collaborative environment for engaging novices in scientific inquiry in phenology; and MoClo Planner – a multi-touch interface for collaborative bio-design. Our evaluation framework consists of three layers, which examine the *usability*, *usefulness*, and *impact* of reality-based interaction on creative problem solving in a collaborative context. Drawing upon the existing body of work in the area, our framework documents a mixed-method approach that aims to provide guidance (rather than an extensive checklist) for the evaluation of reality-based creativity support environments.

Following, we describe our evaluation framework. We then demonstrate its application in the evaluation of G-nome Surfer 2.0.

MULTI-TIERED EVALUATION FRAMEWORK

The proposed framework consists of three layers that examine the *usability*, *usefulness*, and *impact* of reality-based creativity support environments. The first layer applies a micro perspective – focusing on the usability of concrete interaction techniques and the effectiveness of individual visualizations. The second layer applies a macro perspective – studying the usefulness of a system in the context of a full-scale task. Finally, the third layer applies a holistic perspective – examining the impact of the system on users' performance in-situ. Table 1 provides a summary of our evaluation framework. For each layer, we describe its dimensions, settings, metrics, and methods for data collections.

		L1: Usability	L2: Usefulness	L3: Impact
	Dimensions	Functionality Learnability Performance Memorability Errors Satisfaction	Effectiveness Efficiency	Performance Engagement Collaboration
	Settings	In Lab Study	Comparative Study	In-Situ deployment Longitudinal evaluation
	Metrics			
Task	Completion	x	x	x
	Workload	x	x	x
	Accuracy	x	x	x
	Time on task	x		
	Number and type of errors	x		
User	Attitudes	x	x	
	Satisfaction ratings	x		
	Levels of participation		x	
	Equity of participation		x	
	Engagement ratings			x
Learning	Nature of discussion		x	x
	Nature of collaboration		x	x
	Problem solving strategies			x
	Number of hypotheses			x
	Methods			
	Expert review	x		
	Logging information	x		
	Observation	x	x	x
	Discourse analysis		x	x
	Video coding		x	x
	Questionnaires	x	x	x
	Interviews		x	x
	Debrief	x		

Table 1: A multi-tier evaluation framework for collaborative tabletop interaction.

The first layer, *usability* (L1), consists of six dimensions that draw upon Schneiderman's definition of usability [14] and are not specific for reality-based interaction. These include: 1) functionality – the ability of the system to support the user in completing a required set of tasks; 2) learnability – the extent to which it is easy to learn how to use a system; 3) performance – the extent to which the accomplishment of a task satisfies known standards of completeness, accuracy, and speed; 4) memorability – the ability of the user to re-establish proficiency using a system after a period of inactive use; 5) errors – the frequency, type, and severity of errors as well as how easy it is to recover from errors; and 6) satisfaction – the degree to which a user finds the system pleasant to use. These dimensions are easily quantifiable using mostly task-centered metrics as specified in Table 1.

The second layer examines *usefulness* (L2), the advantages of a system for accomplishing creative tasks in a collaborative setting. It consists of two dimensions: 1) effectiveness – the extent to which users' goals are obtained through an effective collaborative process, a process where group members actively communicate with each other to demonstrate shared effort [5]; and 2) efficiency – the degree to which goals are obtained with the investment of less time as well as physical and mental effort. Effectiveness and efficiency are interdependent and should be considered together.

These dimensions can be quantified by combining various task-, user-, and learning-centered metrics that are calculated using mixed methods. For example, the effectiveness of a collaborative tabletop interface can be quantified by measuring task completion rates along with rating the effectiveness of the collaborative process. Collaboration profiles [12] are often useful for describing the nature of collaboration by highlighting the different roles participants assume throughout the collaborative process. Computing the level of participation per user is helpful for calculating the equity of participation [2]. Dialog analysis can provide further insight into the nature of discussion carried by users while working on a task – helping to identify divergent and convergent thinking. For example, such dialog analysis can reveal the time spent on task-related vs. non task-related talk while highlighting insights gained by users [9] and reflective utterances (which indicate analogical thinking). Efficiency can be quantified by measuring task completion time as well as mental and physical effort. Subjective mental and physical effort is often measured using the standard NASA TLX questionnaire [3].

Finally, the third layer focuses on studying the *impact* (L3) of a creativity support system on users' performance and practices in-situ. This layer takes a holistic approach, studying impact on three dimensions: 1) performance – here we consider performance more broadly than in the usability layer, examining not only quantitative task-centered metrics such as time, completion rates, and workload, but also creativity-centered measures that focus on how users apply creative thinking. In particular, we look into how users solve problems in collaborative settings. We suggest utilizing video and discourse analysis to identify behavioral profiles, problem solving strategies, and the number and quality of hypotheses (i.e. alternative solutions) explored by users; 2) engagement – this dimension goes beyond mere user satisfaction to capture the degree of user's interest, emotional involvement, and dynamic interaction. O'Brien et al. (2008) developed a multi-scale measure for user engagement that considers six attributes of engagement: Perceived Usability, Aesthetics, Focused Attention, Felt Involvement, Novelty, and Endurability; and 3) collaboration – the degree and manner to which users collaborate on a task. Here, we consider various metrics that indicate how users collaborate, what roles they assume during the collaborative process, and whether and when they switch their roles. To rate the effectiveness of the collaborative process, we use the rating scheme created by Meier's' et al [7] that considers 5 different dimensions of collaboration: communication, information pulling, coordination, interpersonal relationship, and motivation.

Taken together these three layers consider creative thinking as a multicomponent process that is strongly influenced by group interactions [1].

Following, we describe briefly the application of this framework to the evaluation of G-nome Surfer. A detailed report describing the evaluation process, results, and findings is provided in [11, 12, 13].

EVALUATING G-NOME SURFER

G-nome Surfer [11, 12, 13] is a tabletop user interface for collaborative exploration of genomic information (see Figure 1). It was designed to support hypotheses forming by facilitating collaborative, immediate, and fluid interaction with large amounts of heterogeneous genomic information. G-nome Surfer utilizes multi-touch and tangible interaction techniques to lower the threshold for using advanced bioinformatics tools as well as to provide support for the divergent, convergent, and analogical stages of the creative inquiry process.

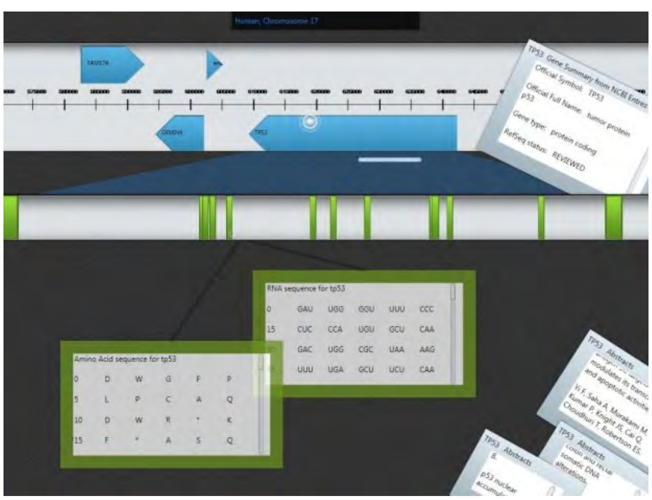

Figure 1: G-nome Surfer 1.0 displaying the human gene TP53 and related publications.

We applied the first layer of evaluation, *usability* (L1), on a continual basis throughout the development of G-nome Surfer. In addition to user testing of each of the complete versions, we often conducted micro-studies examining the usability of particular features through the iterative development and testing of a series of prototypes in increasing fidelity [11].

The second layer, the evaluation of usefulness (L2), was applied through an experimental study with 48 participants that compared undergraduate students' learning of genomics using existing bioinformatics tools (i.e. GUI condition) and two alternative prototypes of G-nome Surfer 2.0: a collaborative multi-mouse GUI and a tabletop interface [12]. We also evaluated the usefulness of G-nome Surfer Pro through a study with 14 student researchers that used the interface for microbiology research.

Our findings highlight several advantages of tabletop interaction for creative problem solving compared to multi-mouse GUI including: 1) *reflective dialogue* (i.e. analogical thinking) – in the tabletop condition, participants spent significantly more time on reflective activities and articulated a larger number of insights than in the other conditions; 2) *physical participation* – participants in the tabletop condition exhibited significantly higher levels of physical participation, expressed by increased spatial manipulation of information. We observed that in the tabletop condition participants manipulated information artifacts – moving, resizing, and rotating – to a greater extent than in the two other conditions. Often, users aligned information artifacts side by side for comparison and then moved them around the table to share with their partner or to place them in an area of the tabletop for later use; 3) *intuitive interaction* – the tabletop condition facilitated more intuitive interaction. This was evident from a statistically significant lower number of utterances related to interaction syntax compared to the two other conditions, and from the reduced time spent on finding information rather than discussing it; and 4) *effective collaboration* – in the tabletop condition, participants were engaged in a more effective collaboration than in the other conditions. This was evident from the turn-taking collaboration style exhibited by most tabletop pairs. Discourse analysis data revealed that in the tabletop condition there were a significantly higher number of coordination utterances and a significantly lower number of disengagement utterances compared to the multi-mouse GUI. Taken together, turn-taking style, higher number of coordination utterances, and lower number of disengagement utterances indicate effective collaboration.

Finally, we applied the third layer of our evaluation framework to study the *impact* (L3) of G-nome Surfer 2.0 in authentic educational settings, deploying it in an intermediate-level undergraduate Neuroscience laboratory course at our institution. Results from this evaluation provide empirical evidence for the feasibility and value of integrating reality-based creativity support environments in college-level science education as well as shed light on how users collaborate and solve problems using such environments in the context of college level inquiry-based learning [13].

Figure 2: G-nome Surfer comparing gene ontology and expression data of different mouse genes.

CONCLUSION

We have presented a multi-tiered evaluation framework for reality-based creativity support environments and described its application in the context of college-level inquiry based learning. The proposed framework consists of three layers that examine the *usability*, *usefulness*, and *impact* of an environment on creative problem solving. This framework takes a holistic approach for gaining an understanding of the strengths and limitations of an environment by utilizing a variety of quantitative measures and qualitative indicators. Combined together, the dimensions and metrics proposed by this framework highlight multiple facets of the creative process mediated by a particular interface.

ACKNOWLEDGMENTS

This work is partially funded by NSF grant IIS-1017693.

REFERENCES

1. Dehaan, R. L. (2009). Teaching Creativity and Inventive Problem Solving in Science. *CBE Life Science Education*, 8(3):172-181.
2. Harris, A., Rick, J., Bonnett, V., Yuill, N., Fleck, R., Marshall, P., et al. (2009). Around the table: Are multiple-touch surfaces better than single-touch for children's collaborative interactions? *ACM Computer-Supported Collaborative Learning* (pp. 335-344). ACM Press.
3. Hart, S., & Staveland, L. (1988). Development of NASA-TLX (Task Load Index) : Results of Empirical and Theoretical Research. *Human Mental Workload* .
4. Jacob, R. J., Girouard, A., Hirshfield, L., Horn, M. S., Shaer, O., Solovey, E. T., et al. (2008). Reality-Based Interaction: A Framework for Post-WIMP Interfaces. *Human Factors in Computing Systems* . Florence: ACM.
5. Kirschner, F., Pass, F., & Kirschner, P. A. (2009). A cognitive load approach to collaborative learning: United brains from complex learning. *Educational Psychology Review*, 31-42.
6. Liu, S., Lu, K., Seifeselassie, N., Grote, C., Francisco, N., Lin, V., et al. (2012). MoClo Planner: Supporting Innovation in Bio-Design through Multitouch Interaction. *Interactive Tabletops & Surfaces.* Cambridge, MA: ACM.
7. Meier, A., Spada, H., & Rummel, N. (2007). A rating system for assesing the quality of computer-supported collaboration processes. *International Journal of Computer-Supported Collaborative Learning*, 63-86.
8. O'Brien, H. L., & Toms, E. (2008). What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology*, 938-955.
9. Saraiya, P., North, C., & Duca, K. (2005). An Insight-based methodology for evaluating bioinformatics visualizations. *Visualization and Computer Graphics*, 443-456.
10. Schneider, B., Strait, M., Muller, L., Elfenbein, S., Shaer, O., & Shen, C. (2012). Phylo-Genie: Engaging Students in Collaborative 'TreeThinking' through Tabletop Techniques. *Human Factors in Computing Systems* . Austin: ACM.
11. Shaer, O., Kol, G., Strait, M., Fan, C., Catherine, G., & Elfenbein, S. (2010). G-nome Surfer: a Tabletop Interface for Collaborative. *Conference on Human Factors in Computing Systems.* 2010: ACM.
12. Shaer, O., Strait, M., Valdes, C., Feng, T., Lintz, M., & Wang, H. (2011). Enhancing Genomic Learning through Tabletop Interaction. *Human Factors in Computing Systems* . Vancouver: ACM.
13. Shaer, O., Strait, M., Valdes, C., Wang, H., Feng, T., Lintz, M., et al. (2012). The design, development, and deployment of a tabletop interface for collaborative exploration of genomic data. *International Journal of Human-Computer Studies*, 746-764.
14. Shneiderman, B., & C., P. (2008). *Designing the User Interface, 5th Ed.* Addison-Wesley.
15. Valdes, C., Ferreirae, M., Feng, T., Wang, H., Tempel, K., Liu, S., et al. (2012). A Collaborative Environment for Engaging Novices in Scientific Inquiry. *Interactice Tabletops and Surfaces.* Cambridge: ACM

Understanding Movement: the Design and Evaluation of Lighting and Wearable Technology in Dance Performances

Renata Sheppard
Henry Luce Scholar
Taiwan National University for the Arts
Taipei, Taiwan
rsheppa2@gmail.com

David A. Shamma
San Francisco, CA
aymans@acm.org

ABSTRACT
In this article, we describe the development and evaluation of a prototype electronic fashion garment to be used in modern dance performances. The objective of creating the garment was to study dynamic lighting in a dancer's movement and eventually to scale to the entire dynamic stage as a creativity environment. In the first prototypes, we designed a classroom project for undergraduate dance students to build their own garments and describe the theoretical mapping between movement and lighting through Laban Movement Analysis. From this, we investigate evaluation methodologies for each stage, the demonstration, the classroom, and the future performances.

Author Keywords
Dance; Modern Dance; Arduino; Lilypad; Fashion; Laban; Movement; Evaluation; Performance; Classroom.

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

General Terms
Human Factors; Design

INTRODUCTION
Dance performances include many components: choreography, lighting, music, direction, and costuming to name a few. Increasingly, as wearable technologies become more prevalent, augmenting dances with various sensors to aid in visualization or sonification is becoming commonplace. In many cases, this augmentation is capriciously utilized in a performance, leaving the choreographer, director, and others to utilize the enhancement as add on effects and not a first order tool for conveying meaning during a performance. In this article, we describe the iterative creation and evaluation of such an electronic garment to investigate how choreographic movement can be enhanced with added technology as a first step towards creating a larger dynamic stage as a larger creativity environment.

For CHI 2013 Evaluation Methods for Creativity Support Environments Workshop. Authors retain all copyrights. **http://ecologylab.net/workshops/creativity/**

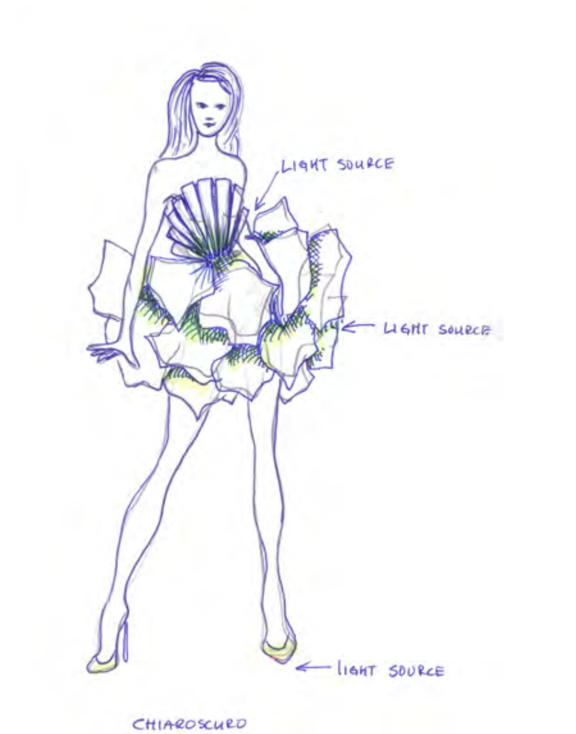

Figure 1. The sketched dress prototype included lighted shoes, dress, and torso.

Our work started with a choreographic question, not a technology one per se. In effect, how would motion effort and quality change if a dancer became illuminated in certain poses? This question drove the technology. However, it was important for the dancer to control the light with his or her body, rather than being controlled remotely by a stagehand or technician; the latter being a proxy for the interaction and hence differently conveyed in choreographic intent and design.

LABAN MOVEMENT ANALYSIS
We used Laban Movement Analysis (LMA) as the basis for codifying movement [3]. It was created through observations of human movement by Rudolph Laban, an early 20th century German choreographer and today has been combined with the work of Irmgard Bartenieffwho researches in movement in the field of physical therapy. LMA is a system of categorizing, notating, observing, and analyzing movement. LMA is a universal framework to describe movement and

provides a bridge between artistic expression and codified descriptions through well-defined concepts about how movement can be identified and categorized. The classification system describes movement with four main categories: Body, Effort, Shape, and Space (BESS). LMA is both a synthesizing and analytical tool that enables very specific awareness and description of movement. Labans practitioners and students often introduce interdisciplinary concepts, including Freudian and Jungian psychological theories, with movement theory. Much of this research introduces new evaluation methodologies, which may seem unconventional, but are increasing in number as researchers begin to surface more interdisciplinary and creative methodologies for evaluation of human-computer interaction, design, and systems research.

DESIGN ITERATIONS

The current stages of this work involve an initial prototype and a classroom experiment. As we are interested in how electronic sensing could map to LMA, we focused on one simple movement and the garment.

Single Dancer

The first prototype started as a simple sketch of the dress and its intended illumination. Figure 1 shows the first chiaroscuro garment prototype. No details were specified at this stage with regards to how the dress would illuminate based on the dancers position. We deferred listing any technical specifications; the initial focus was on the performing aspects of the garment. This includes the visual aesthetic of the dress as well as a choreographic mapping.

For a first step into this investigation, we only wished to explore a single repose. This simplified the electronic circuit design. LilyPad LEDs were used to make a simple circuit with conductive thread [1]; the lights were triggered on and off by a mechanical tilt sensor. This simplified the need for a microcontroller and accelerometer. More so, the tilt sensors position could be physically adjusted and tuned on the dress itself with hot glue. The dress itself was made from artist grade construction paper.

Classroom Experiment

The second prototypes were built from a basic classroom adaptation of the first prototype. The build, materials, and construction were handed to an undergraduate studio dance class with a similar set of build instructions. The students received a priori and in situ instruction on LMA, as well as, hands on construction help from both project members.

The classroom prototypes focused on movement and not on aesthetics. Students used sweatbands and other scraps of garments to make their designs. See Figure 3. The circuit was identical to the first prototype.

Dancer Coordination

The third iteration utilizes a more sophisticated set of electronics. In this embodiment, two garments coordinate an exchange of lighting between themselves. The basic illumination is the same, however, instead of a tilt sensor, an accelerometer is used with a LilyPad [1] Arduino [4] microcontroller. The LilyPad communicates to the other garment over

Figure 2. The first completed chiaroscuro dress prototype.

Figure 3. The two students testing their garments in the studio.

Figure 4. The wireless communicators in prototype 3.

a wireless 802.15.4 XBee serial connection. This prototype, borrowing from the classroom experiment, has an initial technical components and a deferred aesthetic component, see Figure 4.

EVALUATION

Dance and performances have their own evaluation methods both academic and practice, both distinct from Computer Science methodologies. It has been argued that the existing evaluation methodologies in artistic and creative systems should be accounted for in Computer Science when the work is interdisciplinary [6].

The first prototype, being a proof of concept, was submitted to the InLight Richmond festival in 2009[1]. The dress was exhibited with a dancer but no official performance took place. Here we received feedback from peers and curators. More so, lighting conditions at this venue made the garment difficult to see illuminated from afar.

The classroom assignment was intended to challenge the students' creativity in designing choreography and gave us the opportunity to engage with the system and its evaluation. With technology becoming so commonplace in the performing arts, dance programs are experimenting with interdisciplinary courses that require traditional, practice- based dance students to perform and choreograph with technology in order to prepare them for the work field and provide skill sets that match the market. Having had very little personal exposure to technology-based work, this assignment was very challenging for the freshmen studio dance students. See Figure 5.

The lectured framework for the movement assignment was based on the Laban Movement Analysis category of Body which focuses on the overall orchestration or organization of the body and looks at a movement in terms of the initiation, follow through, and resolution. This engages several questions:

- How does the body sequence through the movement?
- What part of the body initiates?
- And how does that part relate to the rest of the body?

The students were told to design a series of movements that simply moved across the floor and in the process triggered the lights to go on and off. Again, the trigger, as in the first prototype, was a simple ball bearing tilt sensor. Surprisingly, even with a very simple mechanic, this proved to be very challenging for the young and inexperienced age group, although a handful of students "got it" with ease. The successful students, who actually completed a prototype of their own, successfully attempted to illicit a reaction from the rest of the classroom, in the form of laughter, smiling, and verbal response, with their prototype. They found a relationship that reflected a cause and effect dynamic between their movement and the goal of turning the light on or off. Each student designed movement that was dependent on their light-costume prototype—collaboration between the movement and the technology. Without the prototype the movement did not elicit the same response and in many cases did not make choreographic sense.

The unsuccessful students created designs and movements with the same impact and significance, even without the costume. This was considered a failure in terms of designing movement, and showed the student's difficulty in relating movement and effort to costume light augmentation.

We then reused several of the students' successful garments in the creation of two wearable costume hats for an informal public showing. See Figure 6. With an understanding of the concept, and having built one themselves, the classroom dancers and they were able to experiment with movement with a more finished product. More so, they continued to find ways to create successful reactions with an audience. The meaning of the costume, the movement, and the dynamic lighting worked symbiotically to create the performance experience.

The process of evaluating a dance typically intersects with the creative process of making the dance. Generally speaking, the evaluation process goes through different levels with three main components: choreographer, performer/performers, and an audience. This happens throughout the creative process from preliminary showings of unfinished work to trusted sources for feedback, much like presenting a prototype to a test user group, to a formal stage production involving lights, costume, theatre and a paying audience, much like launching a beta test product.

In the performing arts, the development process is closely intertwined with the evaluation process. It is important, but not considered necessary, to get feedback from outside sources—peers but not anyone involved in creation, production or performance—continually as part of the process of iteration. Some choreographers may choose not to have external feedback during the creative process for artistic or personal reasons or because they are not seeking a formal or semi-formal evaluation of the process or work. Without focusing on philosophical notions of art making and meaning, it can be generally asserted that the evaluation process in dance relies on a combination of subjective/objective observation of the combined performing elements which include music, sound, costumes, performers, choreography itself, and any additional props that together make up the final composition. The successful performance is the merging construction of these features.

Within conventional academic research, existing evaluation methodologies broadly fall into the qualitative and the quantitative. Towards the latter, there exist opportunities for systems research and evaluation, where advancements in latency, power draw, and lumen performance can traditionally explored (in our case, this would be found in ubiquitous computing research). Towards the former, there exists opportunities to conduct semi-structured interviews with dancers, technologies, and others to explore how the hardware improved the individuals roles and tasks as well as the overall perfor-

[1]Video available at `https://vimeo.com/7315657`

Figure 5. A student prototype illustrating the first ever attempted sewing as well as circuit building for the student.

mance. But it should be noted; the performances success is also evaluated by other metrics like critics reviews and ticket sales.

There still exists the question of evaluating the theoretical mappings between LMA and augmentation technology. We suspect this work, still in its infancy by us, might follow verification and might follow the methodologies of Ekmans work on emotions and facial coding [2] in particular Perlins applications of that work [5].

FUTURE WORK

In this work, we began an exploration in how movement, lighting, and technology could make a dancer want to move? This focused on how a the dancer would "play" with the on/off function as she or he walked around and moved. We aim to scale this interaction slowly, taking time to explore the choreographic possibilities and ultimately gain a larger understanding of the dancers and a stage as a cohesive interactive creative environment. Through the construction of several prototypes, we examined the construction and evaluation of each prototype as either a proof of concept or classroom experimentation—both relied heavily on theoretical grounding and designed to create a choreographic mapping. From the informal public showing, we found that dancers wanted an aesthetic component to the technology as it helps add purpose, intention, and clarity in terms of the movement design. In particular, through iterations, we have found how wearable lighting technology can repeatedly elicit audience reactions. We hope to test the third prototype in a lightweight performance or studio workshop, where we may gleam feedback from the dancers as well. From there, we aim to expand the environment to the entire stage and a set of dancers in a mesh network, connected through choreographic action during a comprehensive performance. Through the creation of these garments, we hope to develop not only a theoretical understanding of movement and technology, but also to codify the mixed methodologies under which these systems can be academically evaluated.

Figure 6. A dance student testing the formalized expansion of his garment. The hat contains the students prototype.

ACKNOWLEDGMENTS

We thank the Taiwan National University for the Arts (TNUA) dance students for their hard work and the Richmond InLight festival and everyone working on the Lilypad and the Arduino.

REFERENCES

1. Buechley, L., Eisenberg, M., Catchen, J., and Crockett, A. The lilypad arduino: using computational textiles to investigate engagement, aesthetics, and diversity in computer science education. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, ACM (New York, NY, USA, 2008), 423–432.
2. Eckman, P. *Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life.* Henry Holt and Company, New York, NY, 2003.
3. Groff, E. Laban movement analysis: Charting the ineffable domain of human movement. *Journal of Physical Education, Recreation, & Dance*, 66 (1995).
4. Mellis, D., Banzi, M., Cuartielles, D., and Igoe, T. Arduino: An open electronic prototyping platform. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (2007).
5. Perlin, K. Creating emotive responsive characters within virtual worlds. In *Virtual Worlds*, Springer (2000), 99–106.
6. Shamma, D. A. *Network Arts: Defining Emotional Interaction in Media Arts and Information Retrieval*. PhD thesis, Northwestern University, Evanston, IL, USA, December 2005.