Title
Interactive Attunement Catalyzes Creative Learning

Permalink
https://escholarship.org/uc/item/3sm491sq

Author
Ngoon, Tricia

Publication Date
2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Interactive Attunement Catalyzes Creative Learning

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Cognitive Science

by

Tricia J. Ngoon

Committee in charge:

   Scott R. Klemmer, Chair
   Steven P. Dow
   William G. Griswold
   Philip J. Guo
   Joy O. Kim
   Caren M. Walker

2021

The dissertation of Tricia J. Ngoon is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

DEDICATION

To my parents. Thank you.

EPIGRAPH

To know, is to know that you know nothing.

That is the meaning of true knowledge.

—Socrates

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Six years and more than ten existential crises later, I've finally finished this thing that we call a PhD. Doing a PhD is not an easy journey, especially during a global pandemic. I certainly could not have achieved this feat without the help of so many amazing people in my life. This is my attempt to show my immense appreciation and gratitude to all those who have led me to this accomplishment.

Thank you to my committee – Scott, Caren, Joy, Steven, Bill, and Philip – for your endless support and feedback. In particular, thank you Scott for being my biggest advocate and pushing me to be a better researcher and thinker. I never would have thought that taking your Interaction Design course on Coursera would lead to me applying to UCSD almost on a whim. I am so grateful that you saw and continue to see potential in me. Thank you to Caren for being such a big influence in my research and a true inspiration. Thank you Joy for being an amazing mentor and welcoming me into the Adobe Research family. Thank you Philip for always checking in and providing me with encouragement. Thank you Steven for the fruitful research discussions and feedback. Thank you Bill for your stimulating discussions and questions during lab meeting and your constant encouragement.

I am blessed to have incredible research collaborators, colleagues, and research assistants. Thank you Adobe Research for giving me opportunities to expand my research, and especially to Mira Dontcheva for all your wonderful feedback and advice through our collaborations. Thank you Michelle Lee, Nicolas La Polla, and Vivian Leung for being such bright and hard-working RAs. This research would not have been possible without your dedication. I am especially proud of Vivian for completing a wonderfully written undergrad thesis on our work together. Thank you Design Lab grad students for creating a supportive and collegial environment within the lab. Thank you to the extended HCI research community and all the people I have had the pleasure of meeting at conferences. In particular, thank you Eunice for being an awesome C&C Student Volunteer Co-Chair and allowing me to share your hotel room when I was stranded

in Austin during CSCW. Thank you Yea-Seul for being a fellow "Rising Star" with me and extensive discussions about our futures after grad school. Thank you Grace for all the insightful conversations interning together at Adobe. Thank you Minsuk for always being a friendly face at conferences. Thank you Jane for all the productive "work" sessions. You have truly been a positive influence on my research and mental health, and I'm a little sad that we won't get to collaborate together at UCSD. Going further back, I am thankful for Tanya Evans's and Christian Batista's mentorship during my time at Stanford. My interest in research all stemmed from Art Shimamura's Learning & Memory class, and I am grateful for his guidance in completing my senior thesis with him.

I am also grateful for the Design Lab and the Cognitive Science Department. Thank you Olga, Sara, Vanessa, Ian, and Michelé for welcoming me into the Design Lab with open arms. Thank you Teenah Eco for being the first person I met in the lab and always being a smiling and friendly presence. Thank you Colleen for frequently checking in on how my family was doing. Thank you to Don Norman for being supportive of everyone in the lab and providing insightful feedback throughout my time here. Thank you Beverley for being the British grandmother I never had and always lending a supportive ear. Thank you Ethel for all your help in getting me to the finish line. Thank you Thanh for supporting me and the rest of the department through all your hard work. Thank you Marta Kutas and Andrea Chiba for helping build my formative research years through my 2nd and 3rd year projects. Thank you Drew Walker for being an advocate for TAs and helping me grow as an instructor. Thank you Doug Nitz and Eran Mukamel for standing up for CogSci grad students when we needed it most.

A huge amount of gratitude goes out to my support crew of friends. Thank you to the crewtons for all the sushiboocha and other shenanigans. Thank you to my tiny, but mighty MARKETS cohort (Michael, Amy, Reina, Kevin, Eric, and Shuai) for having each others' backs all these years. Amy, I remember being roommates with you when we first interviewed at UCSD, and I'm so glad we both ended up here and remained such close friends. Ailie, my two-time

CHAPTER 3, in part, is currently being prepared for submission for publication of the material by Tricia J. Ngoon, Joy O. Kim, and Scott Klemmer. The dissertation author was the primary investigator and author of this material.

CHAPTER 4, in part, includes portions of material as it appears in *Shöwn: Adaptive Conceptual Guidance Aids Example Use in Creative Tasks* by Tricia J. Ngoon, Joy O. Kim, and Scott Klemmer in the Proceedings of the 2021 ACM Conference on Designing Interactive Systems (DIS '21). The dissertation author was the primary investigator and author of this paper.

CHAPTER 5, in part, includes portions of material as it appears in *Interactive Guidance Techniques for Improving Creative Feedback* by Tricia J. Ngoon, C. Ailie Fraser, Ariel S. Weingarten, Mira Dontcheva, and Scott Klemmer in the Proceedings of the 2018 ACM Conference on Human Factors in Computing Systems (CHI '18). The dissertation author was one of the primary investigators and authors of this paper.

CHAPTER 6, in part, is currently being prepared for submission for publication of the material by Tricia J. Ngoon, Vivian Leung, and Caren M. Walker. The dissertation author was the primary investigator and author of this material.

CHAPTER 6, in part, includes portions of material as it appears in *The Dark Side of Satisficing: Setting the Temperature of Creative Thinking* by Tricia J. Ngoon, Caren M. Walker, and Scott Klemmer in the Proceedings of the 2019 ACM Conference on Creativity and Cognition (C&C '19). The dissertation author was the primary investigator and author of this material.

2013	B. A. in Psychology with Honors, University of California Berkeley
2018	M. S. in Cognitive Science, University of California San Diego
2021	Ph. D. in Cognitive Science, University of California San Diego

PUBLICATIONS

Tricia J. Ngoon, Joy O. Kim, & Scott Klemmer. Shöwn: Adaptive Conceptual Guidance Aids Example Use in Creative Tasks. 2021. To appear in *Proceedings of the 2021 Conference on Designing Interactive Systems (DIS '21)*.

C. Ailie Fraser, Tricia J. Ngoon, Mira Dontcheva, and Scott Klemmer. 2019. RePlay: Contextually Presenting Learning Videos Across Software Applications. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, Paper 297, 1–13.

Tricia J. Ngoon, C. Ailie Fraser, Ariel S. Weingarten, Mira Dontcheva, and Scott Klemmer. 2018. Interactive Guidance Techniques for Improving Creative Feedback. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, Paper 55, 1–11.

ABSTRACT OF THE DISSERTATION

Interactive Attunement Catalyzes Creative Learning

by

Tricia J. Ngoon

Doctor of Philosophy in Cognitive Science

University of California San Diego, 2021

Scott R. Klemmer, Chair

Broad exploration is often the best way to solve complex problems because the best solution might not come to mind first. Consequently, creative thinking, like simulated annealing, begins with a "hot" exploratory phase of broad solution search and gradually "cools" to exploitation and narrowing the search space. In solving complex, open-ended problems, because the best approach may not come to mind first, broad exploration is important. However, a common cognitive bias is to "cool" too soon, settling on the first adequate idea that comes to mind or fixating on low-level details, leading to under-exploration of potentially better ideas. The challenge is knowing both what and how to explore, seeing the metaphorical forest for the trees. This dissertation presents how attuning people to a context-appropriate level of detail

at the right time will catalyze creativity. This dissertation addresses three core questions for tackling complex problems: "How do I get started?", "how do I accomplish my goal?", and "how do I move forward?" All three of these questions stymie problem-solving–especially when "better" is multifaceted or ambiguous– and novices especially often fixate on whether an idea is "correct." This dissertation presents interactive system designs that help people surmount these three creative challenges.

This dissertation introduces three exploratory strategies and demonstrates their efficacy. First, abstraction blocks scaffold tangible chunking and visual exploration of ideas to help novices get started in flexibly examining alternatives. Second, adaptive conceptual guidance uses domain-specific heuristics to adaptively present examples depending on a person's progress and task to help them better understand and apply concepts to accomplish their creative goals. Last, interactive feedback reuse combines structural guidance and adaptive suggestions of previously-given expert feedback to help improve critique and move forward on creative work. These strategies and their evaluations demonstrate that structuring exploration and providing adaptive examples in the context of people's work can help people discover otherwise unseen possibilities and better achieve their creative vision.

Chapter 1

Introduction: The Bias Against Exploration Misallocates Creative Attention

On problem-solving, Albert Einstein said, "*If I had an hour to solve a problem I'd spend 55 minutes thinking about the problem and 5 minutes thinking about solutions.*" Yet a common bias in problem-solving is under-allocating up front effort and attention to exploration because exploring a path not taken feels like a waste of time. Why even contemplate a red chair if you know that 'of course' the chair will be blue in the end? The problem with not exploring the possibility of the chair being red is that we neglect the possibility that the red chair might inspire us to do something different with the blue chair, that we will actually prefer the red chair, or that the red chair might be a stepping stone to another unconsidered option altogether. When we lock ourselves into believing that one solution is the only solution, we get stuck on individual 'trees' rather than seeing the full 'forest.'

1.1 Creative Thinking as Annealing

Creative problems are complex; they consist of multiple possible approaches and solutions. Like simulated annealing, creative thinking starts with a "hot" exploratory phase of searching broadly and then gradually "cools" to an exploitation phase of searching narrowly and refining ideas [120, 141, 152] (Figure 1.1). The main insight is that the early stages of problem-solving should consist of high entropy, bouncing around from idea to idea across a diverse spectrum of domains while the later stages should consist of low entropy, focused search. Together, these phases enable broad exploration of possibilities and evaluation of the most contextually-appropriate solution. For example, if you are a designer designing a chair, you first might look across a broad spectrum of chairs from hammock chairs to recliners to office desk chairs before deciding on a simple dining chair and deliberating on details such as color or material. As Schön [179] writes, "This process of experimentation, reflection and action combined, is cyclically conducted as the problem is continuously framed and reframed and as solutions to complex or ambiguous problems are systematically sought." Such effective exploration is difficult because people often satisfice, fixating on the first adequate solution they come up with and not wanting to "waste time" on paths unknown [190]. For the complex problems that make up much of creative work, the cognitive bias to satisfice means that potentially better paths are never explored or considered.

Because of the intrinsic ambiguity of creative work–and the simultaneous co-evolution of problem and solution– scaffolding exploratory thinking is inherently different from scaffolding for problems where a distinctly correct solution exists. The traditional remedy here is personal teaching models like the studio-apprentice model [179] or one-on-one tutoring [21, 34]. However, these resources are prohibitively labor-expensive; they are not always available nor do they scale to the needs and progress of a vast number of novices. There are, however, a few starting points. Intelligent tutoring systems are the most proximal examples of bringing the benefits of

one-on-one tutoring to the masses [14, 125]. These systems work well for problem-solving where the solution and path to the solution are objectively measured. However, in the face of ambiguity, intelligent tutoring systems wilt; tools do not yet flexibly adapt to the contextual aspects of creative thinking. This dissertation introduces interactive scaffolding strategies for amplifying exploratory thinking and reflective dialogue. These strategies and the tools they encompass structure abstract exploration and contextually adapt to a person's progress and activity.
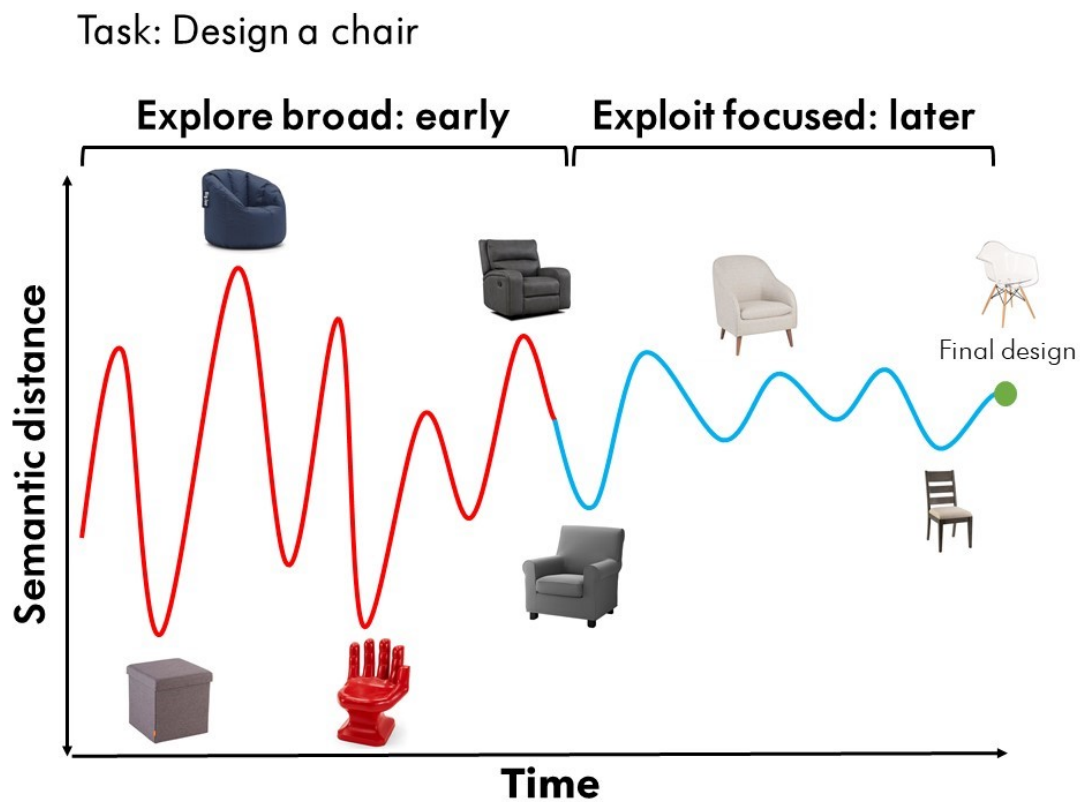
Figure 1.1: Creative Annealing: Like simulated annealing, effective creative exploration begins with "hot" broad exploration, bouncing around ideas with high entropy, followed by a "cool," narrow exploitation phase with less entropy. For example, if the task is to design a chair, one might first start with exploring vastly different types of chairs before narrowing to a more specific type and style of chair.

1.2 The How & the What of Exploration

Exploratory thinking comprises both process (how) and content (what) [111, 112, 152, 159]. For example, a comic artist might first go through the process of brainstorming potential story ideas for their comic ("how do I get started?") and then explore various ways of composing their story idea in their comic ("how do I accomplish my goal?"). Once they have their idea, the artist might seek feedback to help them improve ("how do I move forward?"). This dissertation introduces three strategies–incorporated in interactive tools– that contextually scaffold the process and content of exploratory thinking to help people with the three primary challenges they encounter in creative work (Figure 1.2). First, abstraction blocks scaffold exploratory process by helping people tangibly explore and communicate alternative ideas in visual creative work. Second, adaptive conceptual guidance, embodied in Shown, contextually presents examples based on a user's given progress and activity to improve content exploration. Third, interactive feedback reuse, demonstrated in CritiqueKit, leverages previously-given expert feedback as adaptive suggestions presented alongside interactive structural guidance to also improve content exploration. These strategies and their evaluations demonstrate the potential of structured exploration and adaptive help for amplifying exploratory thinking and overcoming the satisficing bias.

1.3 Structuring Exploration Helps People Get Started in Creative Work

Common models of design process, such as the Britain Design Council's Double Diamond (Figure 2.1) [43] are–for better and for worse–high-level and domain-independent. Empirically, the benefits of such models also remains sparse. Chapter 3 presents a strategy for contextually structuring exploration and evidence of its efficacy.

1.3.1 Abstract Blocking Supports Chunking

Herb Simon's big contribution to our thinking about expertise is two intertwined ideas: 1) expertise is pattern recognition, and experts know more patterns; and 2) Experts 'see' and store patterns in more gestalt, structural terms (what Simon calls 'chunks') [31, 74, 75]. Together, this enables experts to see the metaphorical forest for the trees, while novices who lack pattern-recognition chunks, get stymied by the branches.

Chapter 3 hypothesizes that using tangible chunks provides a malleable way for people to explore compositions for visual creative work. I introduce an 'abstraction blocks' strategy

Figure 1.2: The interactive strategies presented in this dissertation (in bold) scaffold both components of exploratory thinking: the process of exploration (how) and the content of exploration (what). Bolded entries are the strategies in this dissertation that address contextual and interactive scaffolds for both exploratory process and content. Non-bolded entries are examples of other strategies for scaffolding process and content exploration.

that embodies this hypothesis (Figure 3.1). In a collaborative drawing task with six small groups (n=19), groups who used abstraction blocking discussed more high-level concepts and communicated ideas more flexibly than groups who used freeform sketching.

1.4 Contextual Guidance Helps People Accomplish Goals & Move Forward

Creative problems are highly contextual: measured success differs depending on the situation and interpretation. For example, Picasso's abstract drawings are not accurate depictions of the human figure, but are widely regarded as innovative expressions. People need different kinds of help at different times. Assistance that adapts and helps people where they need it most can be powerful for improving creative outcomes. Chapters 4 and 5 utilize adaptive suggestions to provide contextual and individualized help to users at the right moments.

1.4.1 Heuristic Alignment Provides Structure

While creative work is contextual, there are objective heuristics by which creative work is assessed. These heuristics provide necessary structure for understanding the constraints of a problem space. Rubrics are an example of aligning heuristics between a user and the concept, but they require expertise in developing them and do not provide guidance in how to appropriately address rubric criteria. My insight is that helping align people to conceptual heuristics will improve creative outcomes. Chapter 4 introduces Shöwn, which uses heuristic alignment as a way to detect context and provide relevant help for users. In a between-subjects study with novice comic artists, presenting examples based on these heuristics helped people understand why examples were relevant. Chapter 5 applies heuristic alignment as an interactive guidance panel for giving feedback. Two experiments demonstrate that this interactive guidance directs

novice reviewers to give more specific, actionable, and justified feedback.

1.4.2 Adaptive Suggestions Aid Example Use

Expert content in the form of video tutorials or expert-produced examples can help guide people in creating something of their own. These examples are useful for providing a point of comparison and inspiration for novices in accomplishing their goals. However, applying the concept behind an example means understanding when the example is most appropriate and relevant. Chapters 4 and 5 explore using adaptive suggestions to present relevant examples and concepts at appropriate moments depending on the user's task and progress. Chapter 4 introduces the adaptive conceptual guidance strategy and hypothesizes that presenting concepts alongside examples at relevant moments will help novices better understand and apply example concepts. This hypothesis is instantiated in Shöwn, a Wizard-of-Oz system for comic drawing. An experiment shows that adaptive conceptual guidance leads to more unique and readable comics compared to showing a library of static examples (Figure 4.4). Chapter 5 presents the CritiqueKit system, which embodies interactive feedback reuse, presenting previously-given expert feedback as adaptive suggestions. Two experiments demonstrate that adaptively reusing and presenting feedback suggestions (Figure 5.5) improves the quality of critique.

1.5   Thesis Statement & Contributions

My thesis statement is:

Structuring exploration and adaptively presenting help attunes people to a context-appropriate level of detail and leads to better creative outcomes.

My dissertation investigates interactive strategies for attuning people towards exploration in the stages of getting started, accomplishing a goal, and moving forward in creative work.

1.5.1 Contributions

Through deployments and controlled studies of these systems, my dissertation contributes both a theoretical understanding of creative cognition and exploratory thinking and practical applications of how these theories are embedded within creativity support tools.

This dissertation contributes the following:

- Three strategies that structure exploratory thinking to focus attention on higher-level concepts: abstraction blocks (CHAPTER 3), adaptive conceptual guidance (CHAPTER 4), and interactive feedback reuse (CHAPTER 5).

- Interactive tools and techniques that embody these strategies and adaptively scaffold exploration within the context of one's own work.

- Empirical evaluations and observations to demonstrate the efficacy of adaptive exploration strategies on improving creative process and content.

- Observations of exploration as a pedagogical tool for children and designers.

Chapter 2

Related Work: Creative Problem-Solving

Take the problem $2 + 2$. The solution to this problem is true in all cases. There may be a few different ways to solve the problem, such as counting on one's fingers or creating two separate groups of two apples and counting, but the answer is the same and remains the only definitive answer. Now take a problem like decorating a living room. There are principles such as color theory and strategies that designers use, but the outcome of the room itself can vary vastly, and there is not single correct answer to what the room should look like. This is what differentiates creative thinking from other types of problem-solving: creative thinking is one in which both the goal and the solution are open-ended. In problem-solving –creative or otherwise– we engage solution-generating processes of finding all potential paths and verifying processes of choosing among those paths to find a "solution" [152]. The solution-generation process is the "hot" exploratory phase of searching a wide and diverse semantic space and uncovering a breadth of possible ideas; the verifying process then "cools" to exploit, searching a narrower semantic space to refine among potential solutions [120, 141]. The process of exploration (the how) focuses on ideation and thinking about a breadth of alternatives; the content of exploration (the what) focuses on application of ideas and accomplishing goals [111, 112, 152, 159]. Exploration for structured problems where a single optimal solution exists and can be assessed through objective measures

of correctness (*e.g.* solving an arithmetic problem, checking for spelling errors) is straightforward. However, creative, open-ended problems do not have a single or "correct" solution and are assessed based on contextual means other than distinct correctness, making both process and content exploration inherently ambiguous.

2.1   Getting "Unstuck": From a Fixation Bias to Action

People use analogical thinking to handle and manage ambiguity, drawing upon both superficially and structurally similar representations and ideas for problem-solving [64, 65, 69, 206, 213]. A classical example of this is applying the analogy of assembling an army for battle to treating cancer [51, 69]. Analogical thinking forms the basis of creativity, making conceptual leaps through connecting semantically distant, yet relational analogies [29, 66, 67, 82, 215]. People are often drawn to concrete, superficial features because they are more salient. However, focusing too much on the low-level superficial features hinders the ability to see the underlying structure and make further conceptual leaps [64, 69, 141, 215]. For example, in the classical example of applying the analogy of assembling an army for battle, people were less likely to notice the far analogy without a hint because the two scenarios seemed too different, at least on the surface [69].

People often have a cognitive bias to settle too quickly on a single approach when they might be better served by exploring more broadly [101, 113, 157, 190]. Novices especially often lose sight of the forest for the trees, getting mired in details and finding it difficult to take a step back. One reason may be a tendency to engage in fast or automatic thinking rather than effortful thinking to see past details [12, 105]. A person might wonder why they should "waste time" on a path that they do not ultimately pursue. Another reason is an availability bias where the information readily available is the only possible range of outcomes [206]. Lack of expertise is a potential third reason. Unlike novices who get bogged down by the trees, experts are often able to

see the abstract forest, recognizing patterns and creating structural units of these patterns through 'chunking' [31, 33, 55, 74, 75]. When asked to categorize various physics problems, novices categorized them based on superficial features such as whether the problems included a plane or ramp while experts categorized these problems based on their concept such as whether the problem could be solved by the work-energy theorem [32, 33]. Yet expertise is not the panacea for catalyzing exploration. Inflexible, rote expertise might instead be more of an anchor to new and more flexible ways of thinking [95, 96, 225].

Hatano differentiates routine expertise from a more situational form of expertise known as *adaptive expertise* [87]. Adaptive expertise emphasizes transfer of knowledge to new contexts and problems, knowing how and when to apply knowledge through different strategies [180]. In the face of ambiguity as in creative problem-solving, adaptive expertise and analogical thinking function in parallel, prospectively mapping known representations to new and seemingly distant situations [29, 66, 82, 183, 192]. For example, in a study where graduate students and undergraduate students performed a medical diagnosis task, graduate students proactively made external representations that allowed them to perform better at task despite both groups being novices in the medical domain [135]. Expertise alone does not explain the ability to explore, but rather it may be the strategies and the contexts in which expertise is applied that enable greater creativity. In learning situational knowledge and building adaptive expertise, the challenge is orienting people towards an epistemological middle ground of how to find the right design and how to get the design right [26, 202]. The remainder of this chapter will discuss strategies and tools for helping people move away from fixation towards action and exploration.

2.1.1   What are Methods for Catalyzing Exploration?

In any creative project, the first step is coming up with an idea in the first place. Coming up with multiple ideas allows a breadth of options to evaluate [111, 159]. In order to come up with a variety of possible ideas, Osborn [159] presented the rules of brainstorming, encouraging

wild ideas early with the hypothesis that evaluating ideas during generation hinders broad exploration. Process models like the double diamond model [43] also instantiate this hypothesis, distinctly structuring broad divergent thinking from narrow convergent thinking (Figure 2.1). Where Newell's [152] model of problem-solving suggests divergent and convergent thinking in solution-generation and solution-verification processes, the double diamond adds another layer by separating problem-finding from problem-solving. The double diamond gives designers concrete stages of first using divergent thinking in discovering a problem, using convergent thinking in defining the problem and only then moving to solving the problem, once again going through the stages of divergent and convergent thinking.

The creative annealing model does not temporally differentiate explroation and exploitation within problem-finding and problem-solving. More human-centered methods of exploration stretch the boundaries of searching capability further by asking people to think in another's shoes.
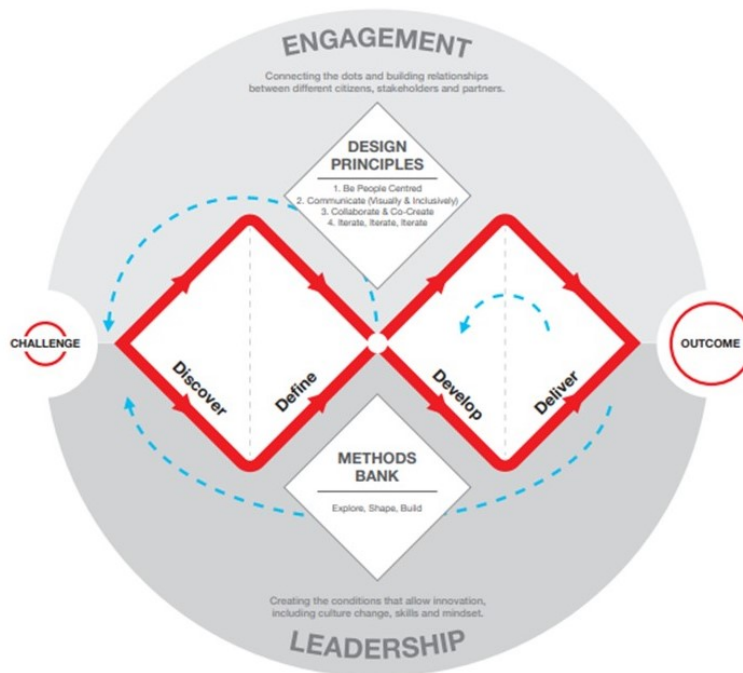
Figure 2.1: The British Design Council created this double diamond model [43] to convey the important distinction between problem-finding and problem-solving, emphasizing them as separate concerns. By contrast, creative annealing is temporally phase-agnostic.

For example, "role-playing" methods encourage people to simulate the "wisdom-of-the-crowd" by taking on the perspective of another person to come up with more diverse and original ideas [42, 41, 46, 199, 217]. Despite explicit prompting, novices and experts alike are still subject to an availability bias in drawing from the most available representation in memory [96, 206]. Exploration is like entering a search query: in order to search for an idea, one must know of its possibility. In other words, ideas are not borne out of nowhere; one must first *see* them.

Two common ways to help people broaden their horizons are examples and iteration. Similar to the idea of self-sourcing, iteration provides a rapid method of exploring alternatives and experimenting with different ideas. Generating multiple prototypes can lead to greater divergence of ideas [49, 48] and improve group dynamics in collaborative creative contexts [47]. In parallel with creative thinking as simulated annealing, iteration enables more exploration rather than maximizing and exploiting on a single solution [49, 47, 190]. Another strategy for promoting exploration is presenting examples. Examples give a comparative structure to beneficially constrain problem-solving [32, 36, 69, 140, 165, 184, 213]. They can also broaden perspectives in making conceptual leaps during creative idea generation [29, 66, 145, 192].

There are, however, limitations to how and when iteration and examples are most useful for creativity. Iteration alone does not necessarily mean greater novelty or divergence as the path to least resistance is incrementally iterating upon the same idea rather than trying a new concept [48, 49, 139]. The mere presence of examples also is not necessarily sufficient for novel creativity. Though examples can spark new ideas, they can also lead to conformity as people may anchor to surface-level features and inherently constrain themselves to a narrow set of possibilities [101, 127, 145, 192]. Abstracting away these alluring details might improve how creative problem-solvers explore alternatives and apply them to their own work [25]. Abstraction blocks (Chapter 3) extends the chunking hypothesis and presents a flexible method for novices to quickly iterate and explore high-level concepts.

2.2 Examples & Tutorials Help Accomplish Concrete Goals

Once an idea or goal is formed, the next step is to figure out a path to accomplishing that goal. Examples can serve both as inspiration for idea generation as well as a comparative roadmap to a final goal. People seek out video tutorials and expert demonstrations to show step-by-step procedures for completing concrete tasks [38, 93, 211, 212]. Examples and tutorials give a concrete way to accomplish a goal. Worked examples in learning show expert problem-solving processes for novices to emulate in their own problem-solving in domains like math and statistics can help learners learn how to solve specific problems [16, 36, 205, 220]. Yet examples and tutorials can be rigid, only showing one way to complete a task. For domains like math and statistics, learning from worked examples can limit transfer to novel problems [73, 184, 185]. Novices often replicate examples because of a difficulty in extracting relevance and applying the structure or concept of examples in new contexts [102].

2.2.1 Examples as Reflective Dialogue

Schön [179] presents an architecture studio scenario where a teacher guides a student how to reframe the given problem, using example sketches to demonstrate key concepts tailored to where the student needs the greatest assistance. The expert in this scenario uses examples to support concepts that fit the novice's situation and where they need the most help. Presenting the right concepts and supporting examples at the right time can lead to a reflective dialogue between a creator and their work in which different alternatives are explored and evaluated for their relative benefits. Similarly, human tutoring studies show how tutors use interactive dialogue to identify knowledge gaps and tailor using examples to help learners better understand problems [34, 149]. Adaptively using examples as a form of support can not only help novices solve a specific problem, but also understand the conceptual reasoning underlying the example.

2.2.2 Providing Help Where & When It's Needed

Reaching a goal in creative work has multiple possible paths. For example, if one wants to take a photo of a sunset, they might decide on what camera angle works best, what the photo composition should be, or how to adjust the lighting to capture the atmosphere they want. Each decision involves considering (sometimes subconciously or implicitly) alternatives to see what works best for the given goal. Learning these contextual skills of creative work often takes place in small group or one-on-one environments such as design studios where people can learn through experimentation and doing [123]. Personalized teaching models such as the studio-apprentice model and tutoring are so effective because help can be tailored to the individual learner, and context is captured through dialogue, question-asking, and deictic communication [21, 34, 37, 76, 78, 79, 123, 179]. However, this personalized help is difficult to scale to the vast majority of novices compared to experts.

Advances in intelligent tutoring systems address the challenge of scale, making conversational learning and continuous assessment of learning available to a large number of learners [8, 14, 80, 125, 173]. Intelligent tutors seek to emulate the adaptive and dialogic assistance a human might provide. These systems provide on-demand help for learners, but help-seeking is a metacognitive skill that requires an understanding of what is known and what is unknown [58]. Novices are often unaware of what help they need and when [5, 10, 78, 79]. Even when on-demand help is available, learners can misuse help features in order to simply obtain the correct solution to problems [18] or engage in unproductive help-seeking behaviors [6, 9]. A second issue in intelligent tutoring systems is the ability to model and predict problem-solving and help-seeking behavior. In domains where problems can be well-structured with objective and defined solutions and solution paths, intelligent tutoring systems can model problem-solving process and solutions, but for problems where the solution and metrics of success are ill-defined, intelligent tutoring systems currently lack to flexibility to adapt the help and guidance a learner might need. For one, intelligent tutoring systems are labor-intensive to build and require extensive

domain and programming knowledge [227]. Example-tracing models that learn generalized problem-solving behavior rather than specific procedural models to increase the efficiency of building an intelligent tutoring system and the efficacy of the skills tutored [7]. However, intelligent tutoring systems are still primarily used and studied in more structured domains such as mathematics, statistics, or physics.

Adaptive, contextual guidance can give people help where they need it. Computationally-generated guidance can present examples tailored to a user's current task or creative intent [62, 107, 130, 134]. Some tools include contextual help through community-driven and crowdsourced annotations where previous users can add their own insights in common confusion areas, typically in tutorial videos and search tools [39, 116, 118, 147, 222]. Other tools embed assistance that adapts in reaction to user prompts and queries,showing relevant examples, tutorials, or inspiration depending on the user's task [39, 59, 61, 62, 133]. However, user-directed approaches depend on the user knowing what help they need or a precise search query [6, 10, 20, 78, 91]. The open question is how might one ask for help if they don't know what to ask for?

Two exemplar systems take a proactive approach for both visual and text-based work. CritiqueKit (Chapter 5) uses past expert critique as examples and suggests these examples as a reviewer gives feedback based on whether the feedback fits certain criteria [153]. This approach uses text-based context to determine whether feedback fits structural characteristics to adapt what type of examples are most helpful for a novice reviewer. In visual work, adaptive templates for photography suggest gridlines to help novice photographers apply composition concepts to their photos, using visual context as a basis for suggesting composition options [100]. These proactive emulate human expert-novice exchanges where human experts proactively suggest help by inferring context from a novice's actions and dialogue and guiding a novice towards developing their own explanations and solutions [35, 37, 79, 179]. A technology-mediated proactive approach suggests examples or guidance that a user might not otherwise know or consider [84, 137, 147]. We extend computational guidance approaches by adapting not only the examples shown but also

the *concepts* they embody to help novices see past the details and better understand and apply the insights within examples. Adaptive conceptual guidance (Chapter 4) and interactive feedback reuse (Chapter 5) are proactive and adaptive strategies that present examples in-situ to help users explore different ideas and better understand how to apply relevant examples.

2.3   Moving Forward with Feedback

Once an idea is brought to initial fruition, the next step is receiving feedback to improve and move forward. Because creative work is inherently ambiguous, feedback is critical for improvement and learning to "close the gap" between where a person starts and their ultimate goal [68, 88, 89, 155, 174]. One problem with exploiting too soon is that if work is already in its completed or final stages, feedback may not give sufficient time or opportunity to reach a desired goal [89, 155]. Especially in the early stages of creative work, feedback aids exploration, giving opportunities to freely test alternatives and course-correcting without the need to drastically change in-progress work [48, 202]. Good feedback, in general, pertains to a specific part of the work being critiqued, provides a concrete and actionable suggestion for improvement, and provides reasoning behind why the suggestion is being given so the receiver understands the underlying principle of the suggestion [68, 89, 155, 174, 194, 195]. However, giving feedback is not an explicitly taught skill, and even experts are often unaware of when their feedback is ineffective [96, 155, 194].

Typically, feedback occurs in learning environments such as the classroom or design studio. With the proliferation of digital tools and massive open online classes, the number of learners exponentially outweighs the number of experts or instructors available [110, 233]. Video tutorials and online video lectures can scale teaching of concepts, but lack the ability to scale giving individualized feedback to all learners. With the increased demand for feedback compared to the number of experts available, two strategies emerge. The first is through clustering and

categorization to improve the efficiency of propagating expert feedback [24, 90, 106, 154, 162, 191]. This strategy often works well for domains where errors and classifications are easily determined and categorized such as programming [86, 90, 191]. Yet the burden still remains on experts to provide personalized feedback for more subjective and ambiguous domains like design. The second strategy is peer feedback where peers themselves give critique to one another can reduce this labor burden on experts [128, 175, 203, 204]. Peer feedback provides a more scalable solution for open-ended domains, but peers tend to view this feedback as less valuable than expert feedback [40, 70, 229]. Unlike experts, peers tend to focus on low-level surface features, suggesting improvements to incremental details (*i.e.* suggesting a specific font size on a website) [83, 94, 126, 232]. While these details are important to the overall finished goal, they can take focus away from higher-level conceptual details that are more relevant early on in creative work and hinder further exploration [94, 128, 129, 232].

2.3.1   Improving Novice Feedback

To help novices attend to higher-level critique features, templates or rubrics provide a general structure to follow. Templates and rubrics give reviewers key criteria to attend to when giving feedback, improving feedback quality [17, 19, 27, 108, 126, 142, 232]. Other forms of structural scaffolding include comparing artifacts or examples [27, 108] and prompts to calibrate or reframe feedback [94, 129]. These examples of structural scaffolds anchor attention to high-level features to improve the focus of given critique. They provide pieces of expertise that novices would otherwise not know. Rubrics and templates are generally inflexible, meaning the user must adapt their feedback to the template rather than the template adapting to the user. Dynamic checklists and rubrics give some level of adaptivity [19], but these methods still rely on expert curation to generate templates or criteria and novices to figure out how best to utilize the scaffold.

Reusing feedback as suggestions is one way to overcome the expert curation issue. Suggesting previously-given feedback as examples can help novice reviewers notice features

they would not have otherwise while also being an efficient way of providing feedback as reviewers do not need to come up with feedback on their own [83, 128, 142]. Commercial tools like Gradescope [2] and Turnitin [3] allow reviewers to reuse their own feedback on multiple submissions. Such reuse is most useful for situations where the same issue recurs, and reviewers have an eye towards reuse. However, feedback reuse is typically focused on efficiency rather than the quality of feedback. Interactive feedback reuse (Chapter 5) combines structural scaffolds and feedback suggestions to leverage previously-given expert feedback and interactively present these suggestions according to the structural heuristics of specific, actionable, and justified feedback to improve novice critqiue.

Chapter 3

Abstraction Blocks: Thinking with Boxes for Flexible Abstraction & Collaboration

Effectively communicating ideas is an important part of creative collaboration, particularly in getting started and generating ideas. One major challenge of collaboration is the pull of concrete details over conceptual concerns as concrete details often stand out more than high-level structure. For example, discussing specific details of a drawing early can pull attention away from assessing whether the drawing's composition works as a whole. This chapter investigates using a strategy called *abstraction blocks*, movable shapes to represent size and position of drawing elements without the need (or ability) to specify concrete details, for collaborative drawing. We hypothesized that structuring abstraction around these blocks enables more flexible exploration and communication. In an observational study of six synchronous small groups ($n = 19$) sketching collaboratively, we found that abstraction blocks enabled more iterative and flexible exploration of drawing compositions. Groups using abstraction blocks explored alternative drawing compositions and discussed high-level concepts such as point-of-view and scale compared to sketching alone.

3.1   Introduction

We communicate with others at all stages of creative work, from ideation to critique [143]. For example, when writing, collaborators often first generate a high-level outline, later shifting focus to a lower level by revising grammar [194]. Visual work like web design also spans

20

abstraction levels, from task goals to deciding specific element placement [124]. Early on, high-level concerns should dominate. For example, in drawing, exploring various compositions early is more important than deciding what specific characters look like. Lightweight sketches serve as exploratory tools to visually communicate ideas in domains from architecture to illustration [26, 207, 209]. Dancers use marking to physically think through choreography without focusing on the details of the dance [122]. Designers use paper prototyping [193] and storyboarding [132] to collaboratively think through a goal without focusing heavily on concrete details. These 'blocking' strategies help people view work as high-level chunks rather than individual elements [31, 75, 33] While abstract chunking and high-level exploration are important for both ideation and collaboration, the emphasis of online collaboration in creative tasks has been on refinement and revision of completed work. Focusing on details early on can shift a group's attention to refinement rather than rough exploration, which can lead to fixation on a single idea [101, 190] or iterating incrementally on a previous idea rather than trying a potentially new and better concept [49, 139, 231].

We know that abstraction helps *individuals* focus on high-level concerns [31]; a combination of early stage abstraction and structure may also help *collaborators* focus on high-level ideation and communicate with one another. We hypothesize that explicitly supporting 'chunks' for early-stage design catalyzes collaborators' ability to revise at an appropriate level of abstraction by breaking down sketches into discrete, adaptable parts. Inspired by expert uses of chunking and tools that use tangible interfaces and techniques [124, 169], this chapter investigates using abstract blocks as a form of visual chunking and a tool for structuring group communication. For example, a drawing of a tourist at a train station might consist of several abstract blocks to represent placement of these elements (Figure 3.1); these blocks can be directly manipulated to communicate desired changes, suggestions, or ideas without the immediate need to change the sketch itself. Users can choose to label the blocks either through text or pictorially, giving flexibility in how they choose to represent their ideas. We focus on sketching as an initial exemplar
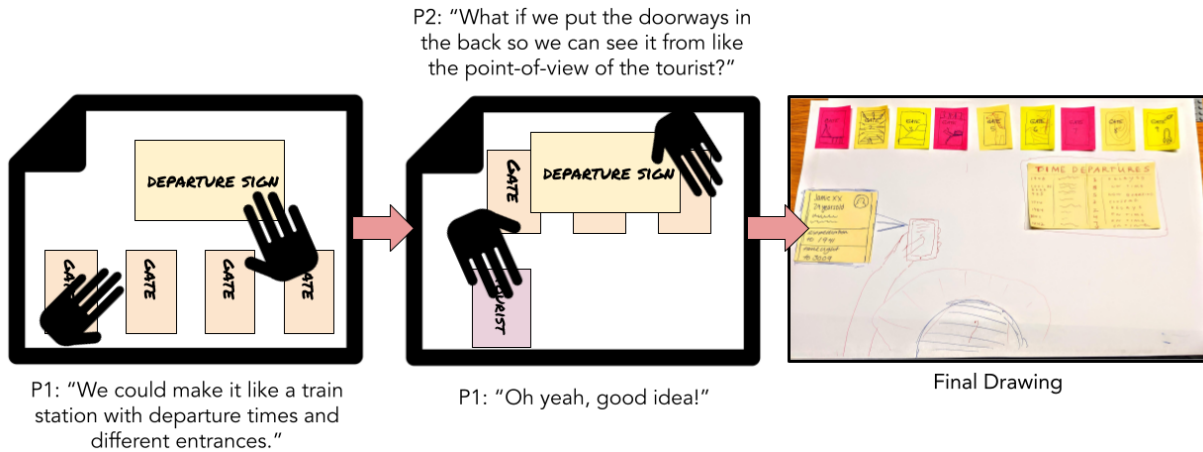
Figure 3.1: A group from our study discussed composition and layout ideas for a collaborative sketch using abstract blocks. These tangible and malleable blocks remove the need for smaller details and emphasize exploration. The prompt for the sketch was "A scene with a time-travelling tourist."

domain of the broader class of creative work because sketching is an easily accessible medium that only requires pen and paper and is a beneficial tool for visual communication [26, 208].

An observational study investigated abstract blocks in a collaborative task. Six groups of novices ($n = 19$) created drawings collaboratively. We audio and video recorded sessions and interviewed groups in a retrospective think-aloud. Groups used a paper implementation of abstraction blocks, using sticky notes, to plan a sketch based on a prompt. Working with simple blocks helped groups to see their drawings in terms of malleable, conceptual chunks. Tangible blocks helped groups discuss high-level concepts like 'point-of-view' or 'scale' and helped them quickly explore and change the composition of the drawing by editing the configuration of the blocks. Working at this more abstract, block level appeared to keep early discussion on more conceptual discussion rather than lower-level implementation details. These results suggest to us that offering (or even enforcing) abstract creation primitives like blocks catalyzes an emphasis on strategic and conceptual issues.

3.2 How is Abstraction Used in Creative Work?

One definition of abstraction is that higher-level features are stable and static whereas lower-level features are those that can change in context [187]. From an information storage perspective, high-level abstractions form units of information, or chunks, formed from recognized patterns [31, 74, 75]. Early in creative work, higher-level abstractions enable the exploration of concepts rather than details such as when dancers use marking to think through an overall choreography [122] or when illustrators or designers use preliminary sketches to communicate rough thoughts [26, 132, 207, 209]. Rubrics and templates are other forms of abstraction tools because they highlight structural goals that are adapted and refined through the addition of concrete details [15, 232]. While experts utilize abstraction techniques like sketching to explore and engage in a reflective dialogue with others and their work [179, 208, 209], novices are often drawn to details [101, 145, 192]. Particularly in visual creative work, this may be because novices are hesitant of their drawing ability or perceive even sketching to be more permanent [93, 223]. This chapter presents abstraction blocks as a strategy for malleable, high-level chunking in visual creative work that eliminates the necessity for early-stage sketching.

3.2.1 Limited Support for Abstraction in Collaborative Creative Tools

Particularly for online collaboration in creative work, tools and strategies for abstraction are limited. Many online projects take a crowdsourcing workflow approach, breaking down larger projects into smaller microtasks with a global goal to ensure coherence even in asychronous collaborations [85]. For text-based projects such as writing a collaborative story, workflows that decompose the work into concrete chunks and allow for interdependent contributions can facilitate structured, high-quality work [117, 119, 170, 176, 210]. Visual creative work like drawing is similar to that of written work in that an overall theme and composition need to be developed to create a cohesive final artifact, but while text has inherent sequential structure, images are not
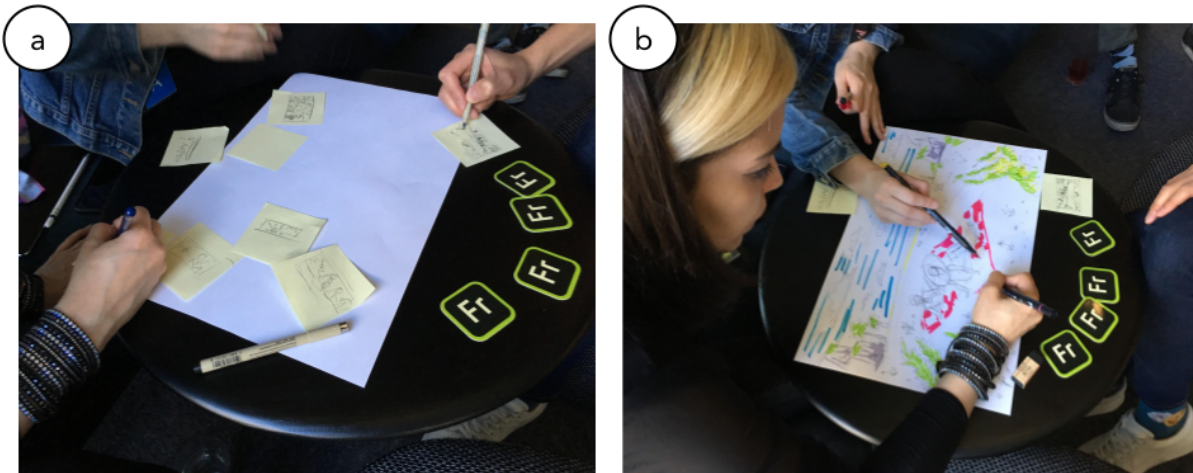
Figure 3.2: a) Experts started planning by sketching multiple thumbnails that conveyed overall structure and composition more than details. b) These composition sketches provided a reference for experts during drawing.

linear and not as easily broken down into independent workable parts [104].

Communication of abstract ideas within these collaborations may rely more on coordinating actions and developing rapport [44, 45], which can be difficult in online settings where groups must rely on technological affordances to effectively communicate [97, 103]. Strategies for abstraction in collaborative visual work are limited in implementation. For example, SwarmSketch [1] allows users to influence the direction of a collaborative drawing by contributing a single ink stroke however they like, but does not provide tools for communication between contributors or structure for how work should be coordinated. As a result, many drawings become a haphazard collection of strokes that no longer match the original goal and are difficult to change [177]. The Johnny Cash Project [201] also predefines the overall project goal of collaboratively making a hand-drawn music video and coordinates contributor work through individual tasks (such as tracing over a designated image); however, it necessarily limits contributions to a single level of detail for ease of coordination, preventing contributors from influencing the direction of the project as a whole.

"Blank page" tools allow for open-ended abstraction, but are often unstructured. The

problem with remaining too abstract or unstructured, especially in collaborative work, is that communication and consensus may get lost in the woods [95, 96]. Multiplayer experiences such as those in Figma (www.figma.com), Google Docs [219], or collaborative drawing canvases (www.drawpile.net) enable a large range collaboration strategies, from allowing anyone to modify anything through pixel-perfect editing, to splitting the canvas or document into sections and assigning collaborators to each part. Digital whiteboards [99] and sketch interfaces [138] similarly give users the freedom to express their thoughts with as much or as little detail as desired. However, these systems leave it to the users to develop their own external collaboration structures. Without prerequisite knowledge of abstraction strategies, novices especially may continue to lean towards specific details, such as the exact specifications of a sketch [223] or precise wording of an essay [194]. Abstraction blocks ground the exploration and communication of abstract ideas to movable boxes to shift focus away from low-level details while giving tangible artifacts with which to interface.

3.3 Abstraction Blocks: Thinking with Boxes

3.3.1 Formative Interviews: How Do Experts Collaborate?

To investigate the role of abstraction in how experts communicate during collaboration, we observed three proficient illustrators recruited from a local drawing and illustration club perform a collaborative drawing task. In this task, the group made a single drawing together based on an open-ended prompt, "Draw a scene of a place that makes you happy." We provided the group with paper, sticky notes, pencils, colored pens and markers, and a single large sheet of paper as the canvas for their drawing. We observed that experts used abstraction in three main ways.

First, experts used abstraction to quickly capture ideas from members of the group. In planning their drawing, the group first created several small thumbnail sketches on sticky notes as members generated ideas for various drawing compositions as part of their planning process

25

Figure 3.3: An expert participant sketched these two thumbnails to illustrate and explore alternate compositions for similar elements.

(Figure 3.2). Creating small, abstract drawings is a common technique for visually generating ideas and solving problems [163, 164]. In this collaborative context, this technique additionally functioned as a way to communicate and confirm thoughts between members of the group, throwing out various "what if" scenarios: *"What if we made like ice cream dogs? I know that's silly, but it conveys happy."* Sometimes a participant would sketch a thumbnail themselves and place it in the center to show the group their idea. Other times, a participant would draw a thumbnail sketch as a different participant was verbally describing their idea, while asking for confirmation that their sketch of the idea was accurate. In a post-interview, the experts mentioned that they learned the thumbnailing technique from their art education and practice.

Second, experts used abstraction to reduce unnecessary work by reusing elements that had already been drawn. For example, Figure 3.3 shows two thumbnail sketches that contain similar elements in different layouts and compositions. To choose a direction for their sketch, the group placed the thumbnail sketches in front of them and discussed to come to a consensus. These sketches conveyed relative positioning of sketch elements rather than specific details of the drawing. The group decided on an overall composition while adding in ideas contributed through

other group members' thumbnails.

Lastly, experts used abstraction as a way to anchor discussions about how to proceed. After deciding on an overall composition, the expert group assigned work to each of its members. However, rather than assigning work based on areas of the canvas paper, the group delegated work based on the elements in the composition they had developed together: "*You draw the background since it's your design, you draw the foliage, and I'll draw the [character]?*" This gave a clear sense in who would draw which part of the sketch and drew from their rough composition plan.

After completing their rough sketch, the experts began adding details. This part of drawing process was much more improvisational; for example, individual members of the group freely colored in areas of the sketch with little or no coordination with other members of the group. In a post-interview, experts noted that the drawing phases did not require as much communication or coordination, since much of the sketch had been previously outlined, reflecting the key concepts of loosely coupled work and maintaining common ground [158].

3.3.2   Abstraction Blocks: Making the Abstract Concrete

From prior work on expert abstraction strategies and our observations of the expert illustrators, we distill these strategies into *abstraction blocks*. We implement these blocks using paper sticky notes as representations of blocking, or chunking, elements that can be shaped, edited, and moved around as desired (Figure 3.1). We chose to characterize abstraction as abstraction blocks because we expected it would be easy for novices to understand and because it mimics common abstraction strategies used by experts in our chosen domain of drawing [163]. While sketches are an abstract representation themselves, novices especially still focus on details rather than exploratory goals during sketching [223]. Encouraging participants to use blocks allows participants to convey rough composition without the need for sketching, emphasizes the spatial relationships between components of a drawing rather than its details, and express a creative vision without worrying about drawing skill or matching the aesthetic style of other collaborators.

27

3.4 Observational Design Study

In this observational design study, we wanted to observe how groups of people with varying levels of expertise in art would work together on a drawing task, and see how collaboration would be affected by usage of abstraction strategies. While the experts we observed used abstract composition sketches to create the "undersketch" of their eventual drawing, we anticipated that most of our participants would not know how to do this.

3.4.1 Method

Participants

Six groups of two to four people ($n = 19$, 11 female, average age $= 28$ years) participated in a collaborative drawing task. Participants were recruited from a technology company and local art school through an email advertisement and signed up for scheduled time slots. Though group members were not required to know each other, at least two people from each group had an existing working relationship. The average self-reported drawing experience level was 1.95 ($SD = 0.78$) out of 4, with 1 being no experience and 4 being expert level. We randomly assigned groups to two conditions: three to *abstraction* and three to *freeform*. Each condition comprised one group of size two, one group of three, and one group of four to explore how communication potentially changes as function of group size. While most groups were co-located, one group in each condition consisted of distributed participants using an online shared canvas (Google Jamboard) and video conferencing (BlueJeans). Observing these distributed groups gave insight on the impact of abstraction on co-located versus remote collaboration. Participants received a $25 USD gift card for their time. All sessions were video and audio recorded with the experimenter taking note of critical moments during each session.

Procedure

For *abstraction* groups, we first presented a short tutorial of how to use abstraction blocks (Figure 3.1). Our tutorial explained that participants could use the blocks to represent where a sketch element would be in a drawing. They were free to use text, sketches, or the blocks as they pleased. Groups each made a collective drawing in response to the prompt, "A scene with a time-traveling tourist." We chose a fictional storytelling prompt and drawing task because it does not require any domain knowledge and enables collaborative ideation and interaction [44]. The task was 20 minutes long and broken into three segments to control for time spent in each part of the process across conditions: five minutes for brainstorming, five minutes for planning and drafting, and ten minutes for drawing. Groups could use paper, pens, pencils, and colored pencils throughout the task. We asked each group to deliver their drawing on one sheet of paper by the end of the study session.

During brainstorming, participants discussed ideas without worrying about drawing yet. Before the planning phase, we gave a short explanation of basic composition concepts and instructed groups to focus on drawing composition during planning. *Abstraction* groups were given standard size and small sticky notes to use and presented with the blocking strategy at this time. Google Jamboard's sticky note feature and pen tools simulated using physical sticky notes and regular analog sketching in the remote sessions. *Freeform* groups were able to plan their drawing in any manner. Following planning, groups collaboratively worked on their drawing for the remainder of the task. Directly after the drawing task, the experimenter replayed recorded critical moments on a computer screen for groups (and through the sharescreen feature for digital groups), asked them to reflect on these clips, and reviewed any additional moments mentioned during the discussion.

3.4.2 Results

To find emergent themes of how groups interacted with drawing artifacts and each other, we noted interesting behaviors and critical moments in each of the sessions, looking for commonalities across groups and between conditions. We found that groups that used abstraction blocks engaged in more conceptual discussion and worked together in a more flexible manner.

Abstraction Enabled Integrated Collaboration

Compared to freeform sketching, abstraction blocks seemed to help groups form a strong shared understanding of each others' ideas. To represent drawing elements, *abstraction* groups placed sticky notes on their draft paper to create a blueprint of their composition. For example, Group 2 (*abstraction*) put three sticky notes together to represent a focal point (a train) of their drawing while using the smaller sticky notes to represent smaller elements (individual people). These sticky notes were labeled with a brief written description or a small sketch showing what the block was supposed to represent.

Similar to the thumbnail sketches created by the experts we observed, abstraction blocks helped anchor discussion by providing shared representations of what members of the group were thinking about: *"I think once someone gave us a visual view of what [they were] thinking, I feel like we all were like 'Yeah! Perfect'."* (P5, *abstraction*). P6 (*abstraction*) similarly said in the post-interview that *"everyone has different ways of placing things on paper so once we laid out that this goes here, I think it was helpful for all of us to just visualize the layout of all the things we were talking about."*

In contrast, Group 1 was the only *freeform* group to make a lightly sketched blueprint of their drawing due to one participant's prior domain knowledge in photography and compositions.

In other *freeform* groups, participants sketched their own individual drawing elements on separate sheets of paper or divided up their shared drawing sheet into areas "owned" by each member of the group (Figure 3.5). For example, Group 4 members each sketched their own

Abstraction

Plans   ### Drawings

Figure 3.4: *Abstraction* groups developed composition plans from the abstract blocks that formed the basis of their final drawings.

Freeform

Plans

Drawings
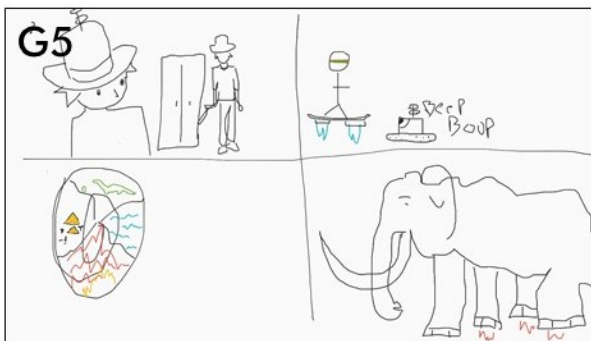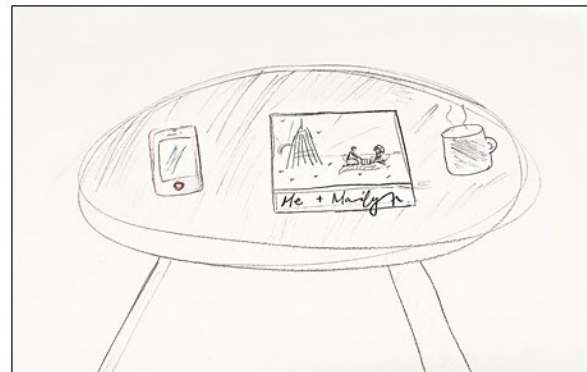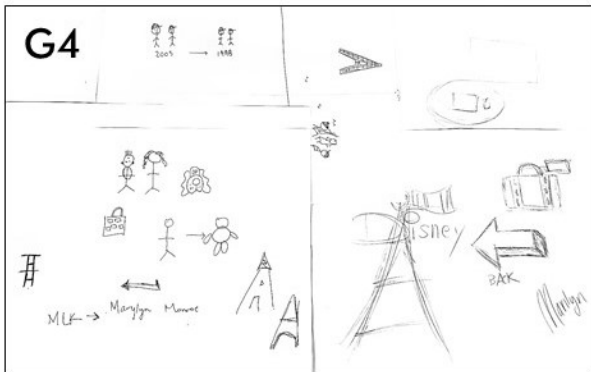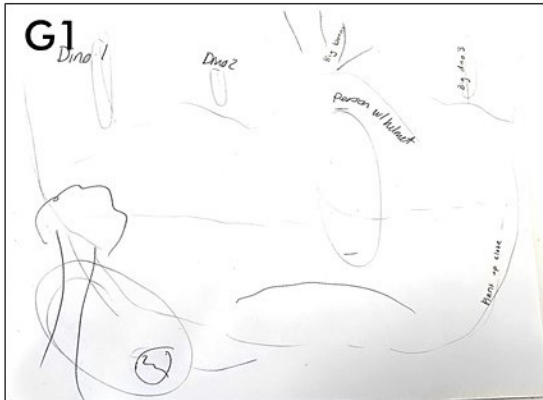
Figure 3.5: *Freeform* groups tended to sketch individual details before deciding on a final composition.

version of the Eiffel Tower on their scratch paper. For *freeform* groups, planning seemed to be oriented around making decisions about specific details of the drawing rather than explore options for the drawing's overall composition.

These differences in how groups approached planning their drawing also affected how they allotted drawing work among individuals. *Abstraction* groups tended to use the blocks they had created as a way to define who should draw what. For example, Group 2 (*abstraction*) assigned each person to draw one of the block elements on their composition plan. They physically checked off the blocks on their plan as they were drawing to mark which ones they completed:

> P3: *"I can take 'colonial' [people]."*
> P5: *"We'll take the 'today' people on this side."*
> — Group 2 (*abstraction*)

Freeform groups focused instead on smaller, individual elements and focused on the details of these elements. For example, Group 1 (*freeform*) split up their work by individual characters: *"So do you want to start drafting the person and play around with outfits, and I'll do the dinosaurs?"* (P1, *freeform*). However, because *freeform* participants had individually planned portions of the drawing separately from others in their group, they encountered conflicts in their understanding of the drawing's composition as sketching progressed.

One *freeform* participant in Group 5 asked another group member to redraw and switch the orientation of another element during drawing: *"Can you draw the portal on the right because I drew the mammoth from right to left, and this will be very complicated for me to draw it from left to right."* (P16 *freeform*). There was also sometimes confusion about what each person should work on during drawing. In the post-interview, P13 (*freeform*) said, *"I didn't quite know what I was doing, everyone seem to have their own parts...and I was like 'what should I do?'"*. P14 (*freeform*) also mentioned that he wished his group had done more concrete outlining beforehand so *"people can work more independently instead of having to wait for another part to be done."*

Abstraction Groups Focused on High-Level Decisions

Groups that used abstract blocks seemed to discuss higher-level concepts like general theme and layouts throughout the drawing process. For example, when talking about how to depict a person looking at a time travel app on his phone, Group 6 (*abstraction*) included light sketches on their blocks (Figure 3.4) to capture and discuss concepts like point of view, background focus, and scale during planning:

> P19: *"Maybe we're still in his point of view, like he's looking at his phone."*
> P17: *"The scale can be small because we probably don't want to get too detailed, but the idea is just to convey that it's a travel app."*
> P18: *"And maybe we can have a thing where it's like zoomed in to show this is what he's holding."*
> — Group 6 (*abstraction*)

Group 3 (*abstraction*), the digital group, spent their planning time setting high-level goals for their drawing. They made blocks with questions for what they wanted to achieve in their drawing rather than making individual sketches: *"Do you wanna have the main question on the top, and then like three vignettes below?"* (P8, *abstraction*)

Freeform groups instead had a tendency to discuss lower-level details during planning, corroborating findings from prior work [101, 139, 231]. In Group 5 (*freeform*), participants discussed incremental details for how specific elements should look. For example, when discussing
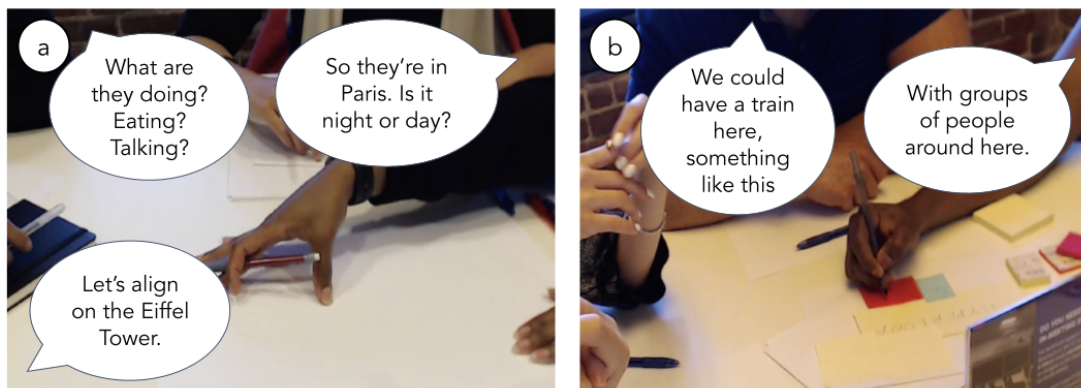
Figure 3.6: (a) *Freeform* groups discussed details early without making concrete sketches until their final drawing. (b) *Abstraction* groups instead used abstract blocks to concretely discuss and sketch a drawing composition.

how to show a time traveller coming out of a time travel portal:

> P15: *"We could have different futures in the portal, like maybe one where the world is flooded."*
> P14: *"Looks like [P15] has gone in pretty hard on [drawing] the portal...so [P13] should draw the traveler."*
> P15: *"And then [P14] will add lasers to it."*
> P14: *"Yeah like laser accessories."*
> — Group 5 (*freeform*)

Discussions would often chain together incrementally, adding more details to existing ideas rather than offering new directions to explore. For example, Group 4 (*freeform*) similarly attended to details and discussed the specifics of what their subjects would be doing in their drawing despite being in the planning stage of the study session:

> P12: *"What are they doing? Talking?"*
> P9: *"Eating?"*
> P11: *"Yeah, wine. Eating."*
> P12: *"Coffee?"*
> P11: *"No, wine."*
> P9: *"Cheese?"*
> P11: *"Yup, cheese. Macarons? Done."*
> — G4 (*freeform*)

In contrast to the expert group we observed, where improvisational discussion of details occurred during the final stages of drawing, *freeform* groups discussed these details during the early stages of ideation despite being encouraged to plan. For *abstraction* groups, structuring abstraction around blocks oriented discussion around the overall composition of their drawing rather than the specific details of how the drawing should eventually look.

Abstraction Enabled Flexible Workflows

Abstraction blocks enabled *abstraction* groups to easily explore alternative layouts and ideas. *Abstraction* groups described drawings as iterative and malleable, using their planning process to ask new "what if" scenarios. For example, while discussing how to draw a train station

for time traveling, Group 6 (*abstraction*) moved blocks to reflect their changing concept and even incorporated the blocks in their final drawing so they could continue to iterate and adapt (Figure 3.4):

> P17: *"What if it's like multiple doors to multiple time periods?"*
> P19: *"Oh and like this [sticky note] can be like the screen that shows all the departures."*
> P18: *"What if it was like this?"* (Takes a sticky note and moves it to another part of the planning canvas)
> — Group 6 (*abstraction*)

Abstraction blocks were useful for planning, but also for making changes during drawing. For instance, during drawing Group six originally drew a screen in the middle of their page, but because they drew the screen on a block, they moved it further to the right side of the page. When asked their reasoning behind this change, P19 (*abstraction*) said, *"There was a moment [when] we realized that we didn't have enough space so that was a moment where we switched things around"*. Abstraction blocks helped in easily adapting their drawing plan even in the later stages of drawing.

P17 (*abstraction*) said the blocks felt impermanent: *"When you draw something, it's permanent, but you can move [blocks] around, so you're not really committed."*. P7 (*abstraction*) also said direct manipulation of blocks helped in getting started: *"Being able to take a sticky note and drag it to where you want it to go, it's so much easier to get started."* (Figure 3.6). Some participants thought the abstract blocks resembled other low-fidelity prototyping methods in helping groups view the drawing as piecemeal rather than a fixed whole. In particular, tangible pieces provided structure in building drawing plans: *"It's about being tangible versus intangible so when we see [a block], we grasp onto the idea of that being something"* P18 (*abstraction*). The flexibility in creating malleable chunks framed the drawing as impermanent, helping groups easily explore different ideas and concepts even during the final drawing phase.

P7 (*abstraction*) also said the abstraction blocks helped in getting started on the drawing: *"it helped in [overcoming] that sort of blank page...and you're so daunted that you don't know*

how to get started...Then when you have a plan you can just get going on the drawing. There wasn't much discussion about what to do in the final stage because it had already been planned out." (P7, *abstraction*).

In contrast, *freeform* groups worked linearly and were often more hesitant to commit to and revise ideas. Sketching, even at a draft stage, was viewed as representing a permanent decision. In the post-interview, P12 (*freeform*) mentioned expressing many ideas verbally, but being hesitant to start their drawing: *"Nothing was concrete on the paper. So it was like we're just spitballing, but once you actually put ink to paper then it's like oh now we need to actually do it."* Group 4 (*freeform*) iterated on specific drawing elements individually (Figure 3.5) because sketches on a shared sheet felt more permanent and committal. They aimed to form verbal consensus rather than exploring ideas the way *abstract* groups did (Figure 3.6). Here they discuss drawing a picture showing the time traveller posing with a historical figure:

P9: *"Do we want to fill the whole page or do we want to do a frame?"*
P11: *"Ooh let's frame it."*
P9: *"So we agreed on a picture, yes?"*
P10: *"A picture within a picture? So is this like a postcard?"*
P9: *"So someone is like holding the postcard-"*
P12: *"Hands. Yes. I like that."*
— Group 4 (*freeform*)

Despite the fact that sketching is often useful as an abstraction tool [26, 209], *freeform* participants in our study still viewed sketching as costly and permanent, akin to final drawing. By providing *abstraction* participants with a way to move their work around, abstract blocks instead made sketching together a more flexible and less daunting process.

3.5   Discussion & Future Work

We investigated how directly manipulable abstract blocks affect communication in collaborative sketching. In an observational design study, we found that abstract blocks gave groups a

way to break down drawing compositions through movable, simple shapes that de-emphasized details and permanence. They allowed *abstraction* groups to communicate higher-level concepts and explore different ideas and compositions relative to *freeform* groups, which used planning phases to discuss more superficial details rather than compositional concepts. In this section, we discuss the implications of these findings for future creativity support tools.

3.5.1   Tools for Supporting Abstract Blocking

This chapter investigated how scaffolding abstraction through abstract blocks affects exploration and communication in the task of collaborative drawing. In any collaborative task, communication can leave ephemeral, intangible artifacts that aid groups in coordinating actions and developing rapport [44, 45]. For example, group dynamic and physical gestures are important for ideation and building consensus in collaborative drawing tasks [22, 198]. In our observations, groups used both verbal and non-verbal forms of communication (like gesturing) to make decisions and share ideas [76]. However, for *abstraction* groups, abstract blocks helped participants concretely express intangible ideas through the placement of directly manipulable shapes, giving groups a malleable visualization of a concept that groups could modify as they conversed. The tangibility of making abstract concepts feel more concrete and importantly, changeable, seemed to be the primary benefit of abstraction blocks.

Groups could, of course, use blocking and similar abstraction strategies without a need for physical blocking tools. However, we observed in our study that *freeform* groups did not use such strategies with the exception of one group who had prior experience in composition concepts through photography. Though sketches are meant to be blocking tools themselves, novices especially do not often separate exploratory sketching from refined drawing [223], as we observed in *freeform* groups. When a concrete abstraction tool—abstract blocks—was made available, *abstraction* groups were more inclined to create abstract representations. Some groups chose to use text to label blocks, others chose to lightly sketch, but importantly, *abstraction*

38

groups did not decorate blocks with heavy detail. Abstract blocks may have helped form common understanding of the drawing among collaborators [158], helping groups maintain discourse on higher-level concepts.

Abstraction through abstract blocks may be even more beneficial in online settings where intangible communication is difficult. In such settings, groups must rely on technological affordances to effectively collaborate [97, 103]. Future collaborative tools could explicitly encourage abstract blocking, as opposed to requiring creators to figure out how to create abstract representations of their work themselves through existing tools (*e.g.*, shape tools in Microsoft PowerPoint or Adobe Photoshop). One might also imagine creativity support tools where abstraction-oriented tools and detail-oriented tools can be used in a complementary manner. Our study also showed that there is a time and place for both kinds of representations in a collaborative creative task, and that there is value in being able to move between abstract and concrete representations of a drawing. Rather than needing two separate creativity tools—one for drafting a prototype and one for generating a final creative work—future work might investigate how a tool could help a group seamlessly navigate between two views of the same work.

3.5.2 Supporting Creative Collaboration

Our studies were done at a relatively small scale; abstract blocks could also support structured, large-scale crowdsourcing workflows. One way people seek to collaborate with a crowd in creative work is through asking for feedback or guidance online through discussion forums or creative communities [131, 186, 221] like Behance (www.behance.com). Much of the exchange and communication occurs via text comment feedback and relies on the creator to interpret suggestions, make changes, and update their work. Methods for structuring communication and exchange among crowd collaborators include assessment [50] and framing [94] prompts, but much of creative work relies on being to *see* and understand changes. In our observations, we saw that collaborators formed a consensus around sketch ideas through abstract blocking. Similar to

bringing in examples as a way to show "something like this" [108], abstract blocks enable quick ways to visually demonstrate ideas without focusing on details.

We also observed that groups used abstract blocks as a way to delegate tasks among collaborators, using the blocks as structured components for each collaborator. At a larger scale, abstract blocking could aid in managing input from a large number of contributors. For example, in leader-facilitated crowdsourcing approaches such as flash teams, a leader could assign modular tasks to a team through blocks [170, 176, 210] or signal which parts of a creative work are open for collaboration [114]. In remixing communities, contributors might use blocking to build upon certain structural components of an existing project [169]. From a social perspective, abstract blocks can also give a tangible way for interactively exploring ideas, which could be particularly useful in creative livestreams where a crowd can engage with an artist and contribute ideas beyond text chat suggestions and audience voting games [60]. Future work should examine how abstract blocking might impact communication at a larger-scale across domains.

3.5.3   Rich Abstraction or Simple Details?

One limitation of using abstract blocks was a decrease in improvisation. Serendipitous creativity can occur through collaborators building and improvising off others' ideas [44, 45]. We saw this in the *freeform* groups where groups would chain ideas together, adding on different details as the group discussed. Because *abstraction* groups created a concrete composition plan ahead of drawing, we observed more expedient drawing, but with less improvisation overall. However, we did observe improvisation and exploration during the planning process when groups were first discussing sketch ideas. *Abstraction* groups explored alternative composition ideas rather than details early on, reflecting the expert process we observed. One area for future work is examining what level of abstraction is appropriate at different stages of creative work. Could abstraction blocks be used to also specify concrete details? This could allow both for exploration of high-level compositions as well as exploration of lower-level details while removing the

need for higher-fidelity sketching. Tools could provide different modes for different levels of abstraction, similar to multi-layered interfaces that adaptively disclose functionality as the user progresses [188].

Another limitation to using abstract blocks is it is unclear *how much* detail should be abstracted. Some groups in our study simply used text labels on blocks to communicate an idea. Others used lightly sketched images. One group used image search to provide a reference for their idea. Each of these representations might differentially impact communication and understanding of the ideas presented. In addition, most *abstraction* groups did not discuss nuanced composition details such as the direction a character is facing or their position in the drawing. Instead, much of the discussion emphasized placement since the abstract blocks focused on adjusting the sketch's overall layout. We saw more of a discussion of these details in the *freeform* group.

Our observational design study presents several opportunities for how tools might implement abstraction strategies such as abstract blocking. The first is in supporting large-scale collaboration. We observed how tangible and malleable blocks helped groups communicate and iterate upon abstract, exploratory ideas. One question for future research is how abstract blocks might support communicating composition details while maintaining flexibility and high-level focus without compromising the benefits afforded by removing detail in the first place. Another opportunity is in comparing the use of abstraction blocks in individuals versus a group and seeing how exploration strategies might change based on the level of communication and collaboration. Lastly, we examined abstraction blocks in sketching, but the concept could apply to any domain where high-level goals can be broken into chunks. Future work should examine the nature of abstraction and how abstraction blocks might be utilized in domains outside of visual work.

Summary

In this chapter, we describe and evaluate using abstraction blocks, a technique where visual content is *abstracted* to simple shapes to facilitate communication and collaboration

on visual creative work such as drawings. By abstracting elements of visual work, we found in an observational design study that people are more likely to explore high-level concepts such as composition and perspective and are less likely to fixate on details of an early idea. Collaborative tasks rely on effective communication between contributors. For novices especially, communicating higher-level goals and changes can be challenging without domain or procedural knowledge. Abstraction blocks scaffold this communication, using simple shapes as a flexible medium to create conceptual chunks rather than detailed sketches. This abstraction makes the invisible visible and the intangible tangible. Collaborative creativity support systems, both for visual creative work and beyond, should include mechanisms for attuning users to the right level of abstraction to explore and communicate higher-level concepts.

Acknowledgements

This chapter, in part, is being prepared for submission for publication by Tricia J. Ngoon, Joy O. Kim, and Scott Klemmer. The dissertation author was the primary investigator and author of this material.

Chapter 4

Shöwn: Adaptive Conceptual Guidance Aids Example Use

Examples are powerful tools for creativity and can provide inspiration and structure. However, novices often don't know when to seek examples or how to apply them. Showing real-time examples targeted to the current task (adaptive conceptual guidance) may help novices consider inspiration more often and better implement the ideas that examples illustrate. We explore this in a Wizard-of-Oz system, *Shöwn*, that presents examples based on the user's current activity while drawing a comic strip. *Shöwn*'s design is informed by interviews with novice and expert comic artists (n=18). A between-subjects experiment (n=24) found that adaptive conceptual guidance improved the clarity and uniqueness of drawings and stories compared to non-adaptive examples. Users also found the examples more useful and inspirational than those without adaptive guidance. Our results present an initial exploration of the challenges and benefits of contextually showing examples in interactive creativity tools.

4.1 Harnessing the Power of Examples in Creative Work

A perennial challenge for novices is figuring out how to get started on a creative endeavor. Viewing and adapting others' work can help novices find traction in a mountain of possibility. Good examples can give inspiration, unveil new ideas, and highlight the advantages of alternative paths [29, 189]. Examples can also help people compare and contrast different ideas across

examples, drawing attention to important features (such as the composition or color scheme used by a painting) that a creator might apply to their own work [11, 32, 67, 127, 145].

Today, with the proliferation of online communities and galleries, real examples of creative work are highly accessible. Communities like Behance (www.behance.net) and HitRecord (www.hitrecord.org) present examples in searchable galleries that enable users to browse, modify, and implement ideas from creators with diverse backgrounds and levels of expertise, bolstering exploration and inspiration [108, 135, 134]. Another strategy for supporting novices is to present examples within tools. Tools such as iMovie (www.apple.com/imovie) and Adobe Spark (www.spark.adobe.com) leverage expertise through provided templates or expert patterns. Novices in particular benefit from templates because they can shape their ideas and work to the high-level structure provided by the template [115, 130, 232]. For more procedural tasks, video tutorials serve as accessible examples that give guidance and important factual information for achieving concrete goals [211] such as knitting a tri-colored scarf or sketching a human face. Presenting examples enables people to follow in the footsteps of others, giving a point of comparison and achievement for their own work.

4.1.1   The Challenges of Applying Examples

In order to use examples effectively, users must be able to assess the context and concept of the suggested example being presented. For example, if a novice photographer sees an example of a sunset photo with an off-center composition, they may not understand how the framing or composition affects the overall photo quality. Presenting examples shares similarities to search interfaces where a primary challenge is how to select and present the most relevant example or result [91, 146]. Even when examples are present and available, novices especially may not understand what examples are most relevant or how to implement the ideas within them, instead replicating surface details [102]. Without requisite domain knowledge, discerning how to build upon examples without simply replicating them can be difficult.

While examples can give inspiration, they can also paradoxically constrain creativity. Examples that are too semantically different from the target goal might actually be harmful for ideation [29, 30]. Timing of examples also matters; examples presented too late are no more beneficial for creative outcomes than not viewing examples at all [127, 189]. Additionally, when given specific examples, novices tend to conform to salient surface features of examples and apply them to their own outcomes, causing potential fixation and decreasing novelty of ideas [101, 127, 145, 192]. Similar to work comparing experts and novices, novices tend to focus on the details of examples rather than on the underlying concepts [33]. An important consideration for presenting examples is choosing the right examples at the right moments.

4.1.2   Adaptive Conceptual Guidance: Examples in Context

This chapter investigates *adaptive conceptual guidance*, which presents in-situ examples and helps users apply insights from them. We show high-level domain principles at specific points in time as a user works to suggest relevant options that they may not notice or consider. In contrast to providing a library of examples or providing templates [108, 135, 144, 200], this approach contextualizes examples based on the user's activity to give specific help at the right moments, which we hypothesize will aid example use and creative outcomes.

We investigate this hypothesis through implementing a Wizard-of-Oz prototype called *Shöwn* that uses adaptive conceptual guidance to help novices explore and apply examples in the domain of comic drawing. We choose comics as a domain because it involves multimedia storytelling and does not require special equipment, making it a useful exemplary domain for studying the challenges related to creative decision-making when multiple media (such as writing, drawing, and story pacing) are involved. To unearth concrete novice and expert differences for this domain, we conducted participant observations with nine expert comic artists and nine novices. We found that novices, in contrast to experts, focus on low-level detail when asked to create a comic strip and struggle to explore alternative compositions and story ideas.

45

We evaluated adaptive conceptual guidance in a between-subjects experiment where participants created a comic strip. This experiment ($n = 24$) compared comics created by participants using one of two versions of Shöwn: a) a version that provided examples for users to browse on their own and b) a version that provided adaptive conceptual guidance and examples to illustrate possible ways of framing or composing comic panels. Adaptive conceptual guidance enabled participants to create comics that were rated by readers as having more unique and clear stories and drawings than comics created without adaptive guidance (Figure 4.4). Raters also preferred comics created with adaptive conceptual guidance overall. Participants stated in interviews that guidance inspired them to consider different ways of showing their story, lending support to the hypothesis that adaptive conceptual guidance assists exploration. Most participants without adaptive conceptual guidance either did not view the examples or did not find them useful, further supporting the hypothesis that novices need guidance for knowing how and when to implement examples to their own work.

4.2 Interviews: Expert & Novice Differences in Comic Drawing Process

4.2.1 Method: Interviewing Expert Artists & Novices

Differences in experts and novices have been well-studied in various domains from chess [31] to physics [33]; experts are able to contextually apply different high-level strategies to their work while novices often focus on details. We wanted to understand where novices struggle most in the domain of drawing comics to learn about the types of examples that would be most beneficial to support in an adaptive creativity support tool. We interviewed and observed nine expert artists and nine novices as they performed a comic-drawing task. We chose comics as a domain because most people would likely benefit from examples when engaging in the

intersection of verbal and visual storytelling [4, 53, 148] and because drawing comics requires only pen and paper at a minimum.

We recruited experts from the freelance site Upwork, and novices through social media posts. Experts had formal art training and professional experience in making comic strips. We recruited novices who had a familiarity with comics, but no experience making comics and no formal education or professional experience in art. Over the course of an hour, participants drew a four-panel comic for one of three randomly given prompts:

- PENGUIN: a man befriends a penguin because of shared interests,

- ALIENS: aliens invade Earth during the 2020 global pandemic quarantine, and

- CHARACTER: a woman is upset at her date because he doesn't know who her favorite television or movie character is.

We asked participants to think aloud as they drew and asked questions about their process. After participants finished drawing, we asked them what they found easiest and hardest about making their comics, what the feel is the most important part of comics, and what type of help they feel would be most useful. All interviews took place on the Zoom video conferencing platform, and participants were allowed to draw digitally or use pen and paper. Experts received USD $60 payment for their participation; novices received a $25 gift card.

4.2.2   Results

Experts Focus on Story & Composition

All nine expert artists expressed that story and composition are the most important parts of a comic: "*It's all about composition because if you can't understand how a comic moves, the narrative kind of decomposes*" (Expert 4). Experts stressed that good drawing skills alone do not make a comic interesting, explaining that writing is more critical to a good comic than the
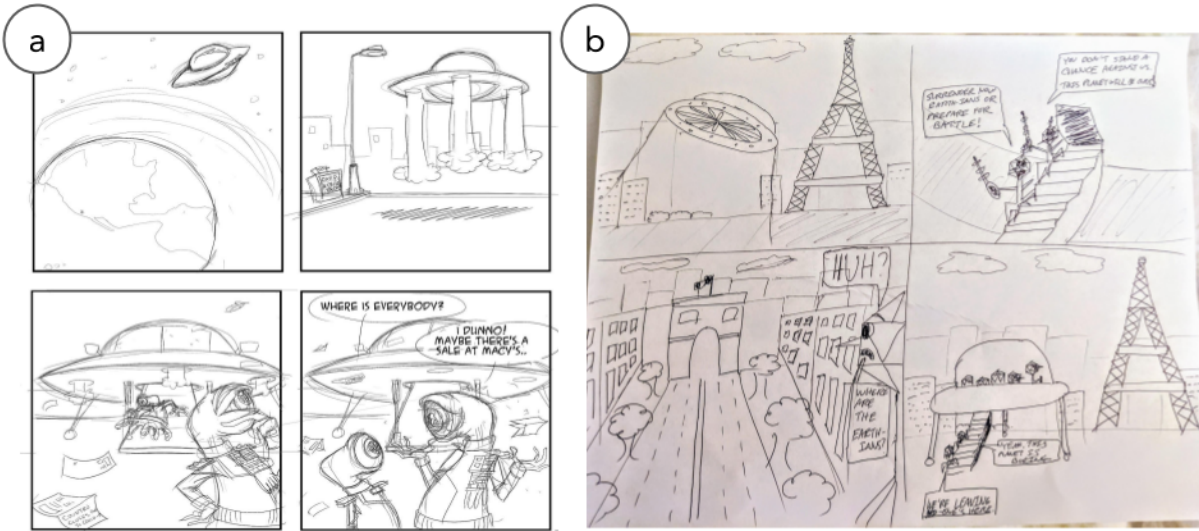
Figure 4.1: Examples of an a) expert and b) novice comic for the ALIENS prompt: aliens invade Earth during the 2020 global pandemic quarantine.

fidelity of its drawings: "*I feel like the most important part is the composition of each panel because you can have great illustration..., but if the composition leads to just a random object in the background that isn't important to the story..., the audience is going to miss something*" (Expert 2). Expert 8 similarly said, "*The most important thing is the writing because if you're a good artist, but your writing stinks, you're not going to make it. But you can have awesome writing and stick figures and make a million bucks, so to speak.*"

To achieve their narrative goals, experts mentioned using composition concepts and rules to help establish the flow and message of the comic. Expert 5 stated, "*There's pretty simple rules of composition, you know, thirds, rhythm, all that stuff. They kind of help lead you to the subject's eyelines.*" As in prior work on expert drawing strategies [52, 164], experts drew rough sketches to rapidly outline their ideas. These quick sketches helped them evaluate their work in terms of visual concepts like composition, as well as narrative concepts like flow and pacing. For example, Expert 6 said, "*So I just plan out how I want it to look. And, you know, once you're done with something you're always like, 'Oh, I can make that better. I can make it nicer.' To get the general gist, we're going to see what they're doing, and etcetera.*" Expert 1 similarly explained, "*And*

at this point, I'm still kind of planning stuff out. So I'm not really drawing what might be there, just kind of a reference. So I guess laying it out." Three experts talked about typically creating thumbnail sketches to ideate panel compositions: "*I'll do like four or five different little sketches or thumbnails of what I think are good for the theme that I chose in my brainstorming session*" (Expert 7). After a rough sketch, experts refine their sketch by adding in details, then move into a final inking stage to finish drawing their comic. Expert 5 said that doing a final drawing was the quickest part of the process because the idea in mind was already on paper. In general, experts spent most of their time on exploration and iteration while focusing on high-level elements like composition and flow, enabling them to finish drawing their comics quickly and simply.

Novices Focus on Details

Verbally, novices also expressed the view that narrative arc is paramount: *e.g.*, "*[The most important thing is] conveying the story I want to tell in as clear a way as possible*" (Novice 8). Despite this, novices time allocation favored detailed drawings over composition and flow. For example, one novice drew the details of a character before deciding on story plot or composition. Another spent nearly 10 minutes drawing a person starting from the feet up in one panel, leading them to rush through subsequent panels: "*I've spent like 10 minutes on this, and I'm only on the feet!*" (Novice 2)(Figure 4.2). In contrast to experts, novices often went straight into drawing their final comic with little conceptual planning or exploration. They were often hesitant to use sketches as exploratory tools and instead wrote out their ideas when conceptually planning their work: "*I know how to communicate my idea through words. That's what I'm familiar with. I'm not as comfortable with drawing*" (Novice 1). Conveying a clear story in the comic was the most important goal for all participants, but novices did not have an understanding of how to show their story visually. Figure 4.1 shows an example of an expert and novice comic drawn for the ALIENS story prompt.

Interestingly, despite novices noting that narrative was the most important aspect of a
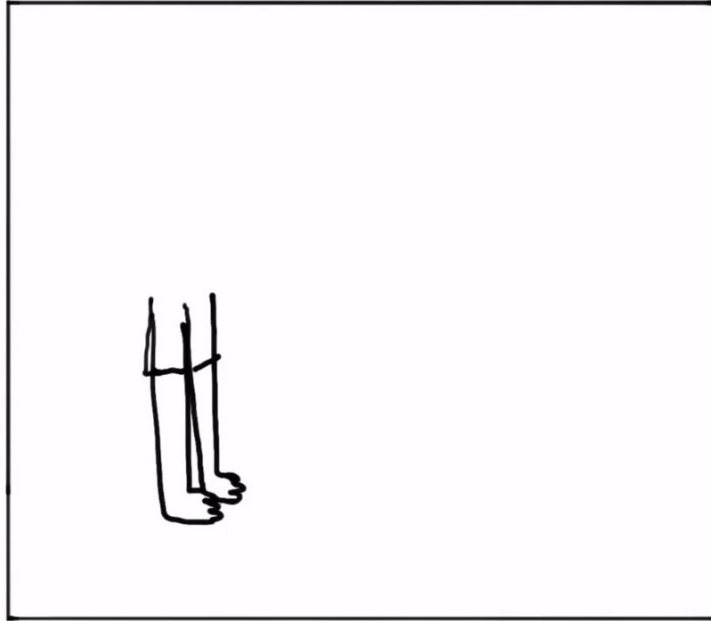
Figure 4.2: This novice from our interviews spent more than 10 minutes drawing a person from the feet up, fixating on getting the details of the character right.

comic, all novices attributed their inability to clearly tell a story to their lack of drawing skills. Novice 2 said, "*I can have an idea in my head, but actually translating it onto a comic is difficult.*" Similarly, Novice 3 expressed that, "*I am able to visualize all of this very easily in my mind's eye but everything else is kind of blank... It's hard to translate that into a specific thing to draw because it's not told to me what that might be.*" When asked what help would be most useful for making their comic, seven of the nine novices said drawing help: "*I think generating the figures is something that could be automated, and [a computer] could have done a much better job than me*" (Novice 7). Another expressed that drawing help would reduce tedium: "*I find drawing to be actually pretty tedious...if I could just come up with the panel and the script and give it someone or the computer that would be great*" (Novice 6). The other two novices expressed a desire for templates or reference images to help them show their comic story.

The perceived lack of skill seemed to be a major factor in how novices chose compositions for their comic panels. Novice 2 said, "*I'm thinking about what I want - like the image in my mind is literally what is on the square. And the way I decide what is on the square is because of my lack*

of drawing ability, it's sort of like what is the simplest way that I can convey the message that I'm trying to set." Exploring beyond the most achievable composition for a panel was also difficult: "*After you've decided on how to present the information or the previous information, and I feel like I don't want to be repetitive..., and I didn't know what would be the best way to draw my panel*" (Novice 4). These observations highlighted two primary challenges for novices: 1) exploring ways to visually show their story idea, and 2) executing on that idea through drawing. Experts are able to quickly evaluate and decide upon alternatives through rough sketching whereas novices often fixate on making their single idea look an precise way.

4.3   Guiding Drawing with Shöwn

Previous research in expert and novice differences find that novices focus on surface details and fixate on ideas [31, 33, 101, 232]. Likewise, our observations showed that novices struggle with exploring alternative ideas and focusing excessively on details while drawing comics. We hypothesize that a system that helps novices consider and choose among options for *what* to draw by presenting timely and relevant concepts and examples can help novices overcome these challenges as they work.

Schön states that reflective practice is a continuous reframing and experimentation of ambiguous problems [179]. Inspired by this articulation of reflective practice, we designed Shöwn to explore our hypothesis. The Shöwn Wizard-of-Oz system augments Google Jamboard, a web-based collaborative drawing tool, with two scaffolds for presenting examples: a drawing helper and adaptive conceptual guidance (Figure 4.3). We chose a Wizard-of-Oz method for two reasons: to study interaction recognition and contextual heuristics for presenting examples without the cost of full implementation, and to better observe how people used and applied conceptual examples [218]. Because Google Jamboard is collaborative, the Wizard and the user can simultaneously see the other's activity by visiting the same link. This allows the Wizard to
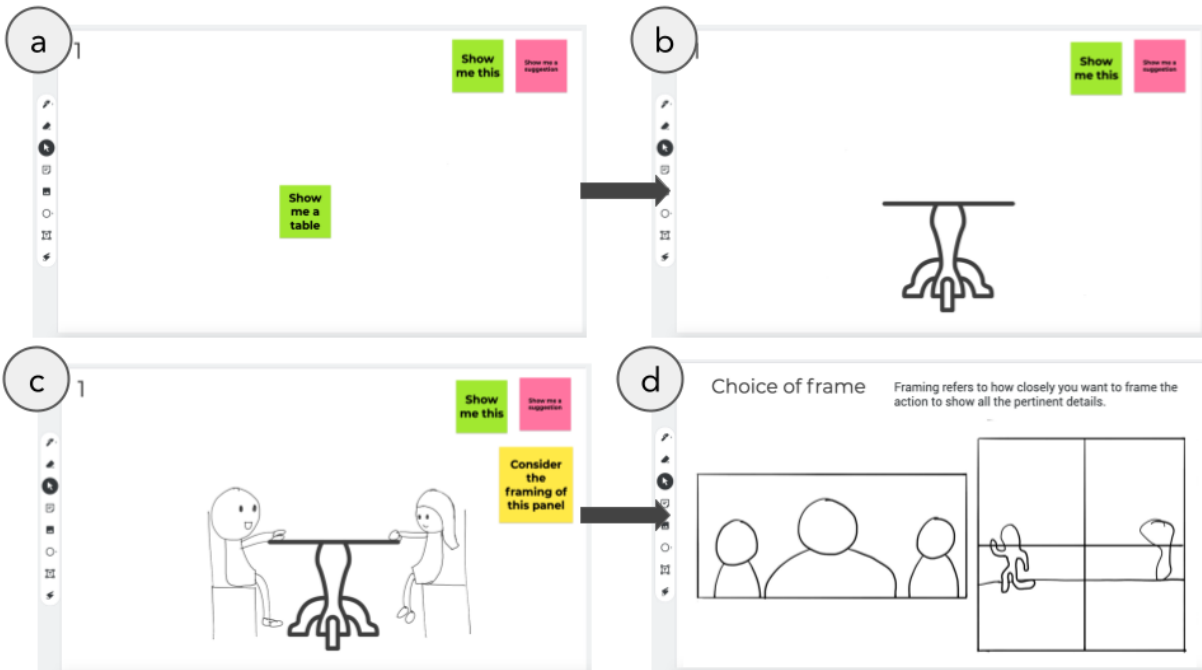
Figure 4.3: Shöwn's interface and features. a) A user wants to draw a comic panel showing a table and requests the drawing helper's assistance by typing a sticky note that says, "Show me a table". b) The drawing helper then provides a basic icon of the requested object in place of the sticky note request. c) As the user draws their panel, Shöwn displays adaptive guidance on the right side of the screen in the form of a sticky note. d) The user can move to the immediate next screen to view relevant examples.

recognize the user's activity and provide assistance accordingly.

The Shöwn interface comprises a drawing area for each comic panel, alongside selected examples. Users can flesh out panels in any order. Following the drawing screens are a corpus of examples, adapted from McCloud's book [148], that show comic-specific concepts related to comic planning such as framing, transitions, and image and word combinations (Figure 4.3d). To communicate with the system, users can use voice commands or Jamboard's sticky notes feature to place notes on drawing screens. Sticky notes as an input mechanism allowed natural language text descriptions and deictic, "put that there" interactions [54].

4.3.1   Adaptive Conceptual Guidance: Aiding Showing Relevant Examples at Relevant Moments

We also found in interviews that, while novices enjoyed coming up with their own ideas and stories, they often did not know how to best visualize their ideas. To address this, Shöwn provides adaptive conceptual guidance by first presenting high-level comic principles as a suggestion and then specific examples as the user works to inspire ideas for *how* to portray their story. The guidance that Shöwn currently provides is based on our interviews with experts and three concepts used by experts to decide how to convey a comic story [4, 53, 148]: choice of framing, choice of moment or transition, and choice of images and words. McCloud [148], a comic expert and artist, writes that these concepts are part of the "*rough planning stage where a story's events are first broken down into readable chunks*" (*pg. 11*). When presenting guidance, Shöwn pairs an explanation of the concept with examples from McCloud's book that show how the concept can be applied to a comic story. (The concept of narrative flow is also mentioned in the book, but this refers to panel layout and spacing, which were not applicable to our study as we asked all participants to use the same four-panel layout.) Highlighting the concept behind examples reflects the scenario of an expert guiding a novice through a problem by explaining high-level concepts and providing specific instantiations of the concepts [179].

The timing of guidance is determined by the following set of recognition heuristics with three principles in mind based on prior work: examples should be related to what the user is currently doing [62, 107, 130, 134, 153], examples should be shown as early in the creative process as possible [127, 189], and examples should be paired with an explanation of the underlying concept they illustrate [102, 153].

The three types of guidance that Shöwn supports are:

Consider the framing for this panel. Framing refers to composition or use of camera angle and distance to show pertinent actions or details in the panel. This guidance is often

presented on the first panel before a user begins drawing to help them get started with their comic. If the user is drawing a subsequent panel, Shöwn presents this guidance if the panel's framing is repeated from any previous panels (*e.g.* if the user begins to use a medium shot in the second panel after using the same camera angle/shot in the first panel).

Consider the moment or transition between panels. Moment or transition refer to the continuity between actions in panels and the specific moments shown as part of the story in each panel. This guidance is presented when a user moves to a new panel to help them think of transitions between panels.

Consider the combination of images and words in this panel. The combination of images and words refers to how words and images together can tell story more powerfully than either alone. This guidance is presented if the images between panels are similar, if a panel contains a lot of or no text or dialogue, or if the user is attempting to draw facial expressions.

4.3.2   Adaptive Conceptual Guidance in Action

While the user is drawing their comic, the Wizard is simultaneously using the same prototype link to provide guidance to the user. As the user is drawing their comic, Shöwn adaptively displays one of the three guidance suggestions based on the user's current task and context. The Wizard displays guidance ambiently [146] on the right side of the panel page without interruption to the user via the sticky notes tool (Figure 4.3c). For instance, a user is drawing the same composition for two panels, in which two people are sitting at a table in the center of the page. Because of the repeated framings, the Wizard presents guidance for the user to "*consider the framing of this panel*" in the form of a sticky note that is added to the right side of the drawing screen (Figure 4.3c) to provide inspiration for alternative ways of composing the panel. When the user sees the suggestion and clicks on the sticky note, the Wizard displays a specific example screen that shows examples corresponding to the concept being presented (Figure 4.3d). The user can then return to their drawing screen and modify their drawing if they choose. When the user

moves to the next panel, the Wizard presents guidance for the user to "*consider the moment or transition between panels*" before the user begins to draw their panel. After clicking on the sticky note, the Wizard presents examples showing different ways to use pacing in a comic narrative. Users can also proactively ask for guidance at any point by clicking the "Show me a suggestion" sticky note on the right side of any drawing screen. Users can also ignore guidance if they choose.

4.3.3   The Drawing Helper: Freeing Focus from Detail

In our interview study, novices experienced difficulty in executing their ideas and overly focused on details due to a lack of confidence in their drawing skills. To mitigate this and enable a larger focus on story, Shöwn provides a drawing helper that can automatically render a basic sketch of a requested object using Google Jamboard's Autodraw feature, which presents icons that match the sketch input. We differentiate these icons from examples presented in adaptive conceptual guidance as these are basic renderings of objects rather than conceptual examples. If the options provided by Autodraw were not sufficient, the human Wizard provided a simple drawn sketch. The drawing helper was designed to draw basic objects such as a table or person rather than complex characters or scenes. For instance, a user wants to draw a comic story about two people having a dinner date. They want to draw the couple sitting at a table, but don't want to focus on the details of the table itself. The user asks the drawing helper for a table by making a sticky note with the text "*show me a table*," and placing it in the center of the drawing screen (Figure 4.3a). The Wizard then produces an icon of the table in the location specified by the user (Figure 4.3b). Later, the user wants a drawing of a chair and this time requests an example via voice command by saying "show me a chair." Shöwn renders an icon of a chair, and the user can move the chair to where they desire on the screen. The drawing helper provides simple drawing objects so the user can then focus their attention on utilizing the conceptual examples that support the high-level narrative aspects of their comic story.

4.4 Experiment

We wanted to test the hypothesis that a creativity support tool that offers adaptive conceptual guidance would help novices better utilize examples and improve creative outcomes than a tool that only provides non-adaptive examples.

4.4.1 Method

Participants

We recruited 24 novice participants (9 male, 15 female, median age = 22) through social media, mailing list advertisements, and the site UserTesting to evaluate our Wizard-of-Oz system. Similar to our interviews, we recruited novices that had familiarity with comics, but no formal education in art or professional experience in making comics. We also required participants to draw using a digital tablet with a stylus for ease of drawing. Half the participants were randomly assigned to the *non-adaptive* condition while the other half was assigned to the *adaptive* condition. The task asked participants to make a rough draft of a four-panel comic based on a randomly-assigned prompt from the interview study (PENGUIN, ALIENS, or CHARACTER prompts). Four participants from each condition saw each of the three prompts. We encouraged black-and-white comics for simplicity, but using color was allowed. All one-hour sessions took place on the Zoom video conferencing platform, and participants received a USD $25 gift card for their participation.

Procedure

At the start of each study session, participants received a link to Shöwn on Google Jamboard and were given a tutorial from the experimenter about Shöwn's features. *Adaptive* condition participants created their comic in a version of Shöwn with adaptive conceptual guidance where conceptual examples were presented selectively based on the heuristics described earlier. *Non-adaptive* condition participants used an otherwise identical version of Shöwn without

adaptive conceptual guidance (the same conceptual examples were available via the example screens, but not selectively presented, similar to previous work with example galleries). All participants had access to the drawing helper because our main research question was about adaptive conceptual guidance versus non-adaptive examples. We chose to include the drawing helper for both conditions because Google Jamboard's Autodraw feature, which is the engine behind the drawing helper's icons, already exists within the tool itself. Participants spent up to an hour drawing their comic while using a think-aloud protocol [56]. At the end of the study session, participants answered post-interview questions about their process and thoughts on Shöwn's scaffolds. Following the study, we debriefed participants to inform them about the Wizard-of-Oz nature of the study.

Measures

We measured participants' behaviors and interactions with Shöwn's features, such as the number of times participants used the drawing helper, viewed examples, and verbally expressed implementing concepts from the examples. We determined whether participants implemented examples either by their verbal expression of doing so or if they changed their panel after viewing examples. We also measured creative outcomes by gathering quality ratings of each comic from lay readers. We chose lay readers as raters (rather than experts) because lay readers are a large audience for comics, and because we wanted to capture overall story impact rather than technical quality, in line with the focus of both experts and novices in our interview study on narrative clarity.

Specifically, we measured comics in terms of clarity, creativity, and overall preference. In pairwise comparisons, raters from Mechanical Turk (61 unique Mechanical Turk workers) were sequentially presented with a pair of comics (two comics for the same prompt, one from each condition). For each pair, raters selected which comic they believed had more unique drawings and a more unique story (creativity), which comic had clearer drawings and a clearer

57
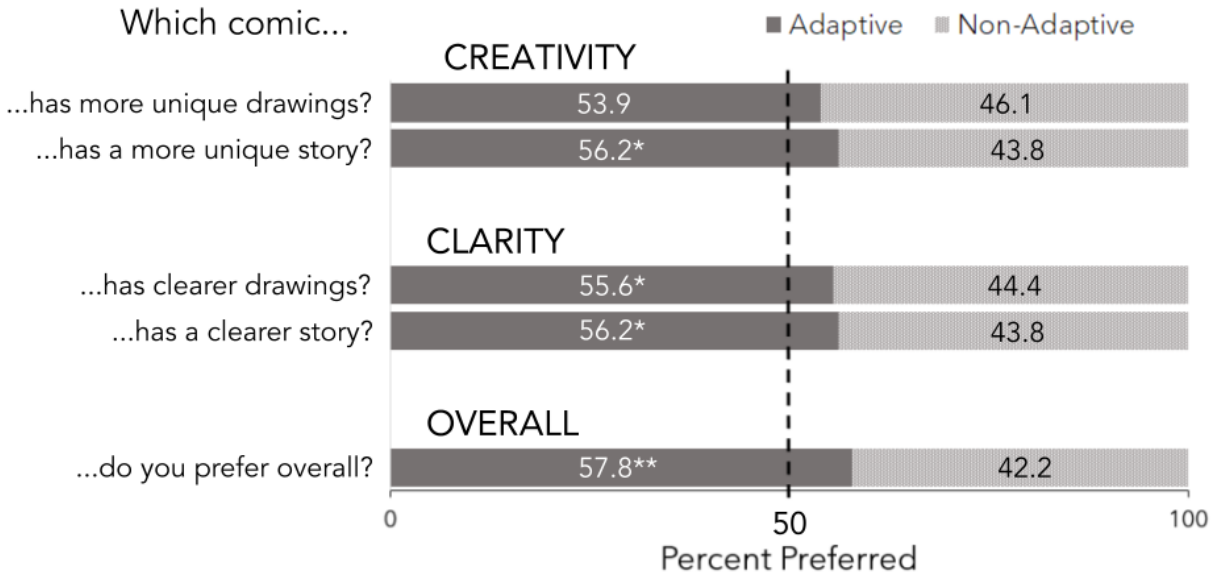
Figure 4.4: *Adaptive* condition ($n = 12$) comics received significantly more preference ratings for more unique stories ($\chi^2 = 5.12, df = 1, p < .05$), clearer drawings($\chi^2 = 3.91, df = 1, p < .05$), clearer story ($\chi^2 = 5.12, df = 1, p < .05$), and overall preference ($\chi^2 = 7.87, df = 1, p < .01$) than *Non-adaptive* condition ($n = 12$) comics. $*p < .05, **p < .01$

story (clarity), and which comic they preferred overall. We separated drawing measures from story measures to account for potential individual differences in drawing ability. Pairing one comic from each condition for each prompt yielded 48 unique pairs (four *non-adaptive* comics $\times$ four *adaptive* comics $\times$ three prompts). Workers were paid USD \$1 to rate five pairs in sequence. Every pair received evaluations from at least five different raters, resulting in 306 total ratings.

4.4.2 Results

Raters Preferred Adaptive Condition Comics Overall and for Most Other Measures

Raters preferred *adaptive* comics significantly more often than *non-adaptive* comics overall (*adaptive*=57.8%, *non-adaptive*=42.2%, $\chi^2 = 7.87, df = 1, p < .01$) (Figure 4.4). In terms of clarity, raters preferred *adaptive* comics more often when asked which had clearer drawings (*adaptive*=55.6%, *non-adaptive*=44.4%, $\chi^2 = 3.91, df = 1, p < .05$) and which had clearer stories (*adaptive*=56.2%, *non-adaptive*=43.8%, $\chi^2 = 5.12, df = 1, p < .05$). *Adaptive*

comics were also more often preferred when raters were asked which had more unique stories (*adaptive*=56.2%, *non-adaptive*=43.8%, $\chi^2 = 5.12, df = 1, p < .05$). There was no significant difference in preference ratings for which comic had more unique drawings (*adaptive*=53.9%, *non-adaptive*=46.1%, $\chi^2 = 2.19, df = 1, p = .14$). Figure 4.5 shows the most preferred comic for each measure as rated by the MTurk workers.

Adaptive Conceptual Guidance Made Examples More Useful

Shöwn presented guidance at least three times to each *adaptive* participant per session with a maximum of six times ($M = 3.92$, $SD = 0.90$). All *adaptive* participants viewed at least one example screen at least one time while drawing, with an average of three times per session ($SD = 1.13$). *Adaptive* participants verbally expressed implementing concepts from examples an average of 2.17 times per session ($SD = 1.11$). In contrast, three *non-adaptive* condition participants viewed examples, with an average of 0.67 examples viewed per session ($SD = 1.23$)($t = 4.84, df = 21.8, p < .001$).

Adaptive participants mentioned that guidance served as inspiration, helping them think of new ways of showing their story they would not have otherwise realized. For example, P5 (*adaptive*) noted, "*As I was going through I think I got into the same rhythm of what I wanted to show, and I think the [guidance] made me more cognizant of ways that I could switch things up. Like I don't think I would have thought of this last panel if I didn't get a nudge from those.*" P16 (*adaptive*), whose comic was the most preferred overall and for the most unique story, similarly stated, "*I think the [guidance] helped me consider different things. For example, for the 3rd panel, I wouldn't have considered doing a shot from the back of his head to show [the woman's] point of view*" (Figure 4.6). For the *non-adaptive* participants who did view the examples, they also served as inspiration. P21 (*non-adaptive*), whose comic was most preferred for having unique drawings, also cited the examples as helpful for their comic: "*[The examples] were kind of like a guide to give me a vocabulary for my intentions.*"
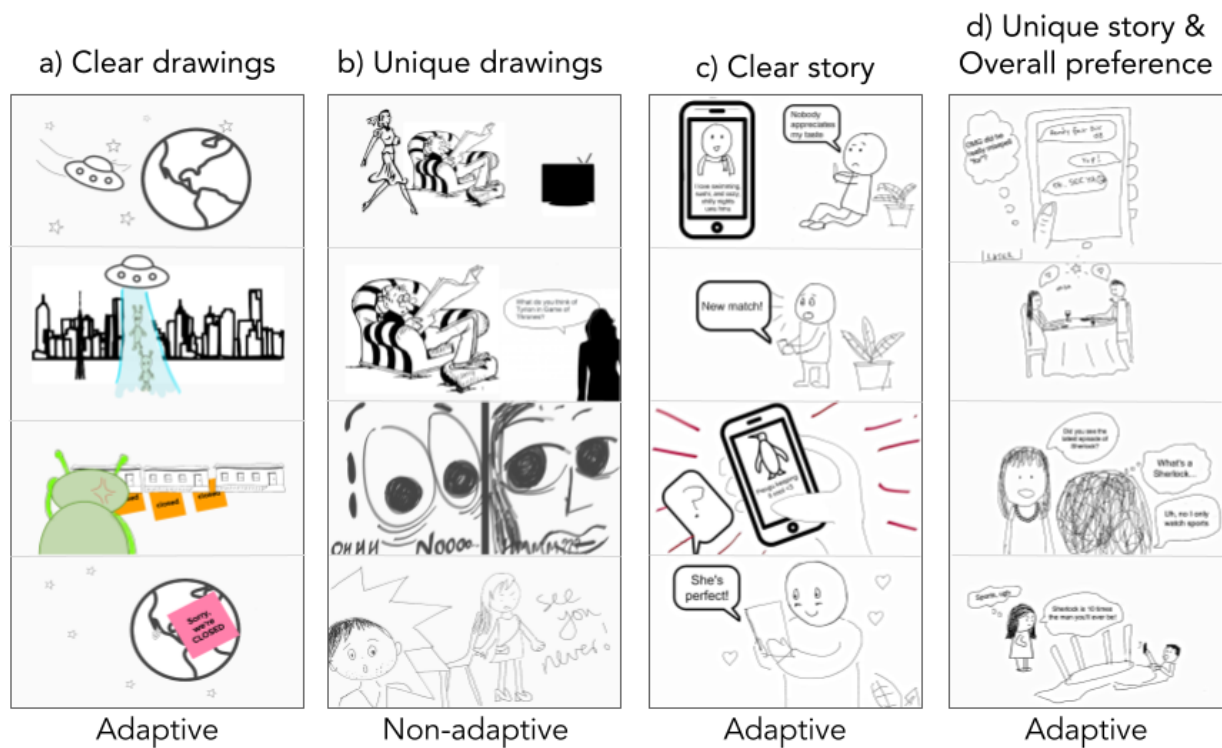
Figure 4.5: The most preferred comics for each measure of clear drawings, unique drawings, clear story, unique story, and overall preference.

However, some *non-adaptive* participants did not believe the examples were applicable. For example, P18 (*non-adaptive*) stated, "*I already had an idea in mind so I didn't really need the examples.*" P15 (*non-adaptive*) forgot the examples were available but wished he had used them after finishing his comic: "*I think I should have looked at [the examples] before I started drawing.*" Similarly, although participants could proactively ask for guidance in the *adaptive* condition, only two participants did so. One participant mentioned that he "*just wanted to see what suggestions were available*" (P14, *adaptive*).

Participants Wanted More Understandable Heuristics for Adaptive Conceptual Guidance

The Wizard uses specific heuristics for displaying guidance. However, the heuristics Shöwn used to decide when to show guidance, even if slightly off, could diminish the effectiveness

60

of examples. Most felt the real-time guidance was useful: "*[The suggestions] were like a guidance for me. I don't think I would go look at [the examples], so it was nice that they were shown throughout the drawing part*" (P10, *adaptive*). However, one participant wished guidance was presented at the beginning of drawing: "*I think these come a bit too late for me to use them because they're sparsed in between while I'm drawing. For example, the perspective one, I thought 'oh, I should've used this earlier,' but the information just came a bit too late for me to use*" (P6, *adaptive*). This suggests that guidance heuristics should not be "one-size-fits-all" but rather allow users to adjust them based on perceived usefulness.

Some *adaptive* participants also thought the timing and reasoning for presenting certain guidance was opaque. They wanted to know more about why particular advice was being given at a particular moment. P24 (*adaptive*) stated, "*The suggestions were helpful, but it wasn't clear why that suggestion was being given. Like for the transition one, it could be that my transition isn't the best one I could use for my viewer or it could just be helping me get started so it wasn't clear why it was giving me those suggestions.*" While adaptive conceptual guidance helped participants better understand examples, clarity around what the system is responding to would help users decide between many possible valid uses of those examples.

The Drawing Helper Shifted Focus from Details to Story

We tracked user interactions with the drawing helper to understand how people used this feature and how it possibly interacted with example use. Participants in both conditions used the drawing helper frequently, with an average of 4.67 times ($SD = 2.87$) per session for *adaptive* participants and 5.67 times ($SD = 4.68$) for *non-adaptive* participants ($t = -0.63$, $df = 18.3$, $p = .54$). In some cases, the drawing helper made drawing objects—particularly repeated objects—less tedious. For example, P24 (*adaptive*) used the drawing helper to draw a phone screen that would be consistent across all panels. P11 (*non-adaptive*) felt the drawing helper did indeed help him focus less on details: "*I thought [the drawing helper] was really cool,*

especially for like tables or chairs or to hash out the overall setting so you can just focus on the other stuff." P21 (*non-adaptive*) thought the drawing helper was useful for showing her intended vision: "*I could just tell [the drawing helper] my vision, and it would bring up things that are relevant.*" Our findings suggest that the concreteness provided by tools like the drawing helper may complement the efficacy of strategies like adaptive conceptual guidance.

However, a few participants felt limited in their creativity with the drawing helper. One participant noted an over-reliance on the drawing helper rather than sketching, stating, "*I kinda became too dependent on [the drawing helper]. Normally, I would just draw it out quickly, but I want to see what the system can do...With the tool, I think if the system can do this for me, and it's perfect the first time, I try to see how I can fit my idea to what the system can do*" (P15, *non-adaptive*). Another participant felt that the simplicity of the drawing helper's sketches did not allow them to completely show their vision: "*It was pretty limited in what it could show. Like*
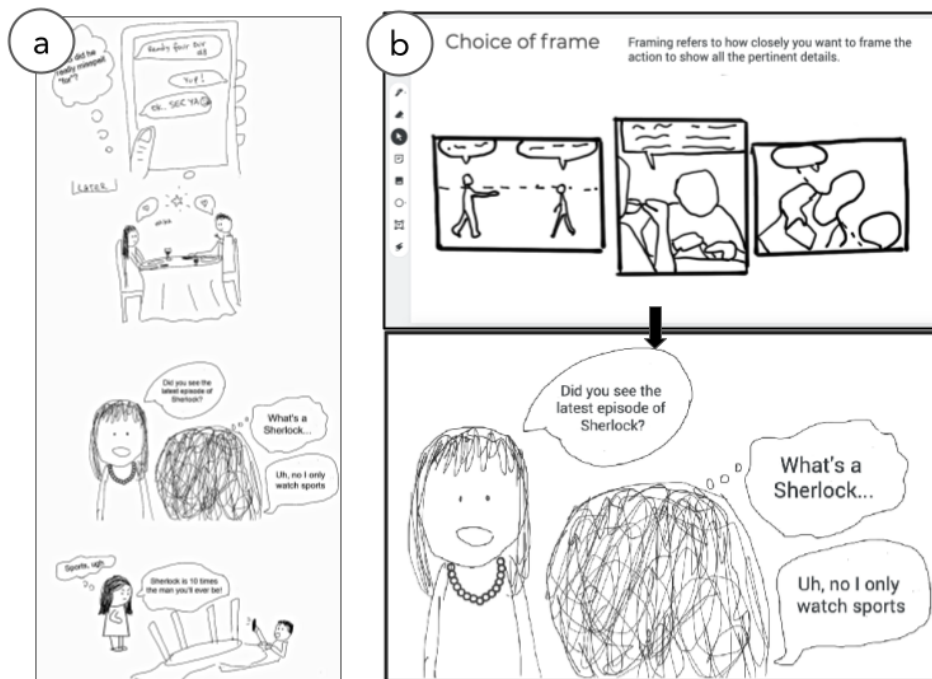
Figure 4.6: a) This comic was created using Shöwn's adaptive conceptual guidance. b) The participant cited that seeing a relevant example inspired the composition of the third panel of the comic.

I couldn't do perspectives or angles with the basic icons I was given" (P12, *adaptive*). These observations corroborate the fact that we found no significant difference in visual creativity across the two conditions ($\chi^2 = 2.20, df = 1, p = .14$). Three participants did not use the drawing helper or only used it once because, as one participant stated, "*It might have been more helpful if I was drawing something more complex, but it was just easier to draw it on my own*" (P13, *non-adaptive*).

4.5   Discussion & Future Work

This chapter investigated whether providing adaptive conceptual guidance aids example use more effectively than non-adaptive examples through a Wizard-of-Oz prototype and empirical study. We next discuss implications of our findings for future creativity support tools and generalization to other domains.

4.5.1   System-Directed Versus User-Directed Creativity Support

In our experiment, we observed that novices in both conditions rarely sought examples without being prompted, perhaps because they were unaware of needing help or did not know what help they needed. Only two *adaptive* condition participants proactively asked for guidance, and only three *non-adaptive* condition participants viewed the examples gallery on their own. While exploring alternatives is a useful problem-solving strategy, meaningfully evaluating them requires expertise and the comparison induces a hefty cognitive load [205]. In line with prior work [20, 78, 189], this suggests that timing is relevant to giving the most appropriate help as novices may not know what type of help is most helpful and when. By incorporating concepts with examples, we found that Shöwn's real-time adaptive conceptual guidance helped to expand participants' field-of-view to concepts they would not have known to explore on their own. Such proactive system-directed approaches seem to benefit open-ended creative problems (like figuring

out how to draw a comic story) because they help users explore beyond known possibilities.

Similar to other systems that present galleries of examples to provide broad inspiration [108, 135, 144], Shöwn provides high-level examples meant for conceptual exploration rather than examples that are particular to the user's specific work. For example, if a user wants to draw two people talking, Shöwn presents general examples of different panel framings or combinations of text and images rather than specific instances of two people talking in a panel. These context-agnostic examples are framed as considerations (rather than statements that impose specific approaches as the "correct" ones). This open-ended approach is analogous to human tutoring, where less didactic interactions allow the learner to form their own explanations and hypotheses in evaluating ideas [37]. In addition, recommending guidance in-situ allowed participants to better understand the relevance of guidance and evaluate examples in context [179]. One participant explicitly mentioned that showing guidance in the moment while drawing encouraged them to view the examples when they would not have otherwise. Participants cited that even if they did not change their initial idea, the provided examples served as inspiration or as a way to help them explain their choices. In these instances adaptive conceptual guidance seemed to serve as a nudge for users to consider different options or evaluate their own decisions.

In contrast, user-directed approaches (such as query-based systems and other help-seeking tools [62, 98, 178]) seem to work best when the user has a concrete idea they want to execute. We saw this with how participants made use of Shöwn's drawing helper, which produced basic icons to quickly visualize ideas based on a user's request. Some used voice commands to direct the drawing helper to provide images. However, the drawing helper had its limits: it could not generate drawings in response to complex, ambiguous requests. For example, P15 (*non-adaptive*) asked the drawing helper to lay out an entire panel: "*Show me a man holding a phone with a dating app on it and lists for interests.*" This participant attempted to shape their ideas around what the drawing helper could generate, rather than using it to show their own ideas. This reflects how novices may form an over-reliance on help-tools or conform to salient features

rather than high-level considerations [101, 102, 145, 192]. Our findings with Shöwn suggest that user-directed guidance works best for small and concrete tasks where the user already has an idea in mind.

Both system-directed and user-directed guidance can support the creative process in complementary ways. Context-agnostic conceptual aid allows for the ambiguity of exploration while specific user-directed guidance can give a concrete direction towards accomplishing goals. In many ways, using examples as guidance is similar to coaching or tutoring, where a coach or tutor gives both open-ended hints or suggestions and concrete examples depending on a person's progress. Shöwn provides a demonstration of using context from user actions to leverage such guidance for creative work. Future research could further examine how computational systems can provide more adaptive coaching or tutoring for creative endeavors.

4.5.2 Designing Adaptive Conceptual Guidance Systems: Limitations & Opportunities

We found two areas where adaptive conceptual guidance could be improved. The first is in determining the timing of guidance. One participant felt that guidance was presented too late and would be more helpful or more relevant earlier. While guidance was shown adaptively, the heuristics for determining when to show certain concepts and examples were static for each participant, which is both a limitation of the Wizard-of-Oz implementation as well as potential technical implementations. One open question is how to better anticipate when to provide guidance when even the user does not know when to ask for help. Potential effective timing mechanisms might be measuring how long a user is idle [28, 189] or taking a mixed-initiative approach in incorporating user input, such as refining or pruning suggestions [107] to determine the best moments to present examples.

The second area of improvement is providing transparency around why guidance was being given. Three participants wanted further explanation for why certain pieces of guidance

were being shown at specific times. One participant asked the experimenter whether the system presented guidance randomly or not. Another thought the presentation of guidance meant the system did not "like" their current drawing. This sentiment reflects prior work showing that people prefer explainable and transparent interactions with AI systems so they not only understand what help is being given but also why [13, 92, 156]. One possibility is to display guidance alongside interactive checkboxes or dynamic rubrics to help users understand how to better situate guidance within the user's context [19, 153].

We also found potential in improving user-directed support. While participants did not often explicitly seek out examples while working on their comics, many participants in both our interviews and experiment asked ideation questions aloud such as "*How do I show this place is empty?*" or "*How do I show that this character is angry?*" as part of the think-aloud protocol. These questions could serve as queries to indicate user intent and aid in tailoring the type of examples shown. During exploratory tasks like brainstorming ideas or deliberating ways to translate an idea visually, users may have more of these open-ended queries in mind rather than concrete goals. Tools that adaptively display help for just-in-time learning as well as incorporate contextual search mechanisms like natural language and deictic instructions may be useful ways of enabling user-directed conceptual support [61, 80, 133, 230].

As a Wizard-of-Oz prototype, Shöwn utilizes a human Wizard to use a user's actions as heuristics to determine what conceptual guidance to provide. Some of the heuristics were based simply on starting on new panels (*i.e.* the Wizard shows the guidance to consider moments or transitions when the user moves on to a new panel). Other heuristics were based on capturing the user's past and current actions (*i.e.* the Wizard presents guidance on images and words if the panel contained too much or too little text). Existing tools already use sketch-based actions as context such as Procreate's Quick Shape tool[1] to automatically complete shapes or straighten a user's lines or Google Jamboard's Autodraw tool that uses object recognition to infer what icons

---

[1]https://procreate.art/handbook/procreate/guides/quickshape//

to suggest to the user. Co-creative intelligent agents can also use object and line recognition to improvise collaborative drawing [45]. Our Wizard-of-Oz evaluation shows how adaptive systems might use visual, sketch-based context to provide conceptual guidance beyond automated drawing help or example galleries. Open remaining questions are how systems might be trained to learn such context as well as what types of context and concepts are most appropriate for adaptive and automated assistance beyond sketching.

One limitation of Shöwn was that the example screens were separate from the drawing screens, requiring users to switch between drawing and example screens in order to view them. Tools incorporating adaptive conceptual guidance or examples presentation should consider more ambient displays that show examples without requiring the user to change the context of their work. Because of the remote nature of the study, one limitation was that we could not control for the screen size participants used. This may have affected drawing ability, though this limitation may have been mitigated since all participants used a stylus and had the drawing helper available.

Tools that help novices overcome skill barriers to execute their creative goals are powerful, but without seeing potential alternatives and possibilities, novices remain bounded by their limited conceptual expertise. Shöwn evaluates simple recognition heuristics through a Wizard-of-Oz implementation to contextually present concepts and examples for better exploration and execution. Future work could examine feasibility and implementation of such heuristics for creativity systems across domains. We find that combining user-led support for concrete queries and adaptive support for open-ended exploration of alternatives helps novices better understand and utilize alternatives for their own work. An interleaving of system-led support and user-led agency is a promising direction for the development and evaluation of future creative tools [92]. We provide a demonstration of this direction that can apply to creative endeavors across domains.

Summary

We present Shöwn, a Wizard-of-Oz system that provides adaptive conceptual guidance by suggesting relevant concepts and examples for a user to explore based on their current task. We hypothesized that adaptive conceptual guidance would guide the application of examples and improve creative work more than simply providing static examples alone. Through interviews and a between-subjects experiment in the domain of comics, we found that adaptive conceptual guidance led to comics with more unique and clear stories and clearer drawings. Users also found adaptive conceptual guidance to be timely and useful for inspiration while users without guidance did not find the examples as relevant and were less likely to take advantage of them. We argue for tools that help novices explore ideas by providing the right kinds of assistance in the right situations. In this way, we may better expand novices' vantage points to explore concepts more broadly and improve creative work. This chapter provides a direction for the future of adaptive creativity support tools and presentation of examples.

Acknowledgements

Chapter 5

CritiqueKit: Interactive Guidance Techniques for Improving Creative Feedback

Good feedback is critical to creativity and learning, yet rare. Many people do not know how to actually provide effective feedback. In providing feedback novices tend to focus on low-level, superficial details, leading to lower quality feedback. This chapter introduces two interactive techniques for improving feedback in creative work: reusable suggestions and adaptive guidance. These techniques are embodied in the CritiqueKit system to help reviewers give specific, actionable, and justified feedback. Two real-world deployment studies and two controlled experiments with CritiqueKit found that adaptively-presented suggestions improve the quality of feedback from novice reviewers. Reviewers also reported that suggestions and guidance helped them describe their thoughts and reminded them to provide effective feedback.

5.1   Introduction: Feedback's Hidden Potential

Feedback is one of the most powerful influences on learning and achievement [89]. Both giving and receiving formative feedback encourage self-reflection and critical thinking on one's work [136, 155], especially in creative and open-ended domains such as design and writing

69

[89, 174]. The growing scale of many educational and professional settings increases both the importance and difficulty of providing sufficiently descriptive and personalized feedback. Good feedback can be hard to generate, and people are not consistently skilled in doing so [128, 232]. Feedback is often too short, vague, and not actionable [129, 194, 228]. Even experienced reviewers don't always recognize when they are providing poor feedback or why it is ineffective [194].

This chapter contributes two interactive techniques that improve feedback in the moment, their embodiment in the CritiqueKit system, and their evaluation through two deployments and two experiments.

Interactive guidance of feedback characteristics. CritiqueKit features a guidance panel with checkboxes that update as the reviewer gives feedback. A text classifier categorizes feedback as Specific, Actionable, and/or Justified as the reviewer types, providing them with an ambient awareness of their feedback quality and guiding them to improve their feedback in real-time while writing.

Suggesting prior feedback for reuse. CritiqueKit enables reviewers to reuse expert feedback, reducing experts' labor by scaling their feedback to similar work. These suggestions update and adapt based on the feedback's categorization to give reviewers targeted ideas for how to improve their comment and provide inspiration.

Two deployment studies and two controlled experiments investigated the efficacy of these interactive techniques on the quality and characteristics of feedback. The first deployment examined how experienced reviewers (teaching assistants) reuse feedback in an undergraduate course. The second deployment examined how undergraduate students reuse feedback. The first experiment examined the impact of statically presented suggestions and interactive guidance on novice feedback. Finally, the second experiment examined the efficacy of adaptively updating suggestions in tandem with interactive guidance on novice feedback. We found that adaptively-presented suggestions improved feedback quality. Reviewers found suggestions useful for

Table 5.1: Two deployments (DEP) and two between-subjects experiments (EXP) examined the efficacy of feedback reuse and interactive guidance. We found that interactive suggestions and guidance were most helpful for improving feedback.

Study	Adaptive Reuse	Static Reuse	Interactive Guidance	Participants	n=	Main Finding
DEP1		X		TAs	8	TAs used suggestions as inspiration
DEP2		X	X	Design Students	29	Students reused vague suggestions
EXP1		X	X	Design Students	40	Static suggestions and guidance were not helpful
EXP2	X		X	General Students	47	Adaptive suggestions and guidance were helpful

inspiration, and the interactive guidance reminded them to ensure their comments met the criteria for effective feedback. This work provides evidence that interactive techniques such as suggestions and guidance can effectively scaffold the feedback process, improving the feedback people give without needing them to spend extensive time learning good techniques (See Table 5.1 for details).

5.2   Related Work

5.2.1   Good Feedback is Actionable, yet Rare

Rapid iteration is critical to the success of creative projects, from essays, to visual design, to buildings [48, 174]. Receiving feedback early on is important for learners to test alternatives and course-correct [48, 202]. Effective feedback is especially important in educational settings where novices are learning new skills and developing expertise. However, giving effective feedback is rarely taught [155]. As physical and digital classrooms increase in size, the demand for feedback outgrows the ability to adopt the ideal learning model of one-to-one feedback [21]. Instead, a one-to-many approach is utilized, where an expert provides feedback for multiple learners. Although learners most value expert feedback [70, 229], the one-to-many approach is

highly demanding on experts, and effective feedback for individuals becomes increasingly rare.

In general, effective feedback is specific, actionable, and justified. Specific feedback is direct and related to a particular part of the work rather than vaguely referent [126, 174, 232]. Specific positive feedback also highlights strengths of the work and provides encouragement, so the recipient can tell they are on a good path [111, 204, 232]. Actionable feedback is important because it offers the learner a concrete step forward [174, 194, 204, 232]. Simply pointing out a problem is not sufficient to help one improve [166, 174, 194, 202]. Actionable feedback is often most helpful early in a project [40, 204, 232] because it may help people self-reflect and self-evaluate their work [68], prompting more revisions for improvement [50, 203]. Lastly, justification is an important characteristic of feedback [126, 151, 232], but is arguably one of the hardest to understand or recognize [70]. Justified feedback contains an explanation or reason for a suggested change, which helps the learner understand why the feedback was given.

5.2.2   Rubrics & Examples Usefully Focus Feedback

Rubrics [15, 232] and comparative examples [126] are effective in structuring feedback because they beneficially encourage attention to deep and diverse criteria. Novices otherwise tend to focus on the first thing they notice, often surface-level details [83, 94, 129, 232]. Viewing examples of past designs can lead to greater creativity and insights [127, 145]; thus, showing examples of good feedback may spark ideas reviewers would not have otherwise considered [83, 128, 142]. Also, adaptive examples curated to match design features are more helpful than random examples in improving creative work [134].

Rubrics and other scaffolds require significant upfront manual work by experts who must carefully design a comprehensive rubric, curate a thorough set of examples, or decide how else to structure the feedback process. This chapter investigates leveraging existing feedback to dynamically create rubric criteria. We hypothesize that showing reviewers previously-provided feedback can guide their attention to important aspects of the design.

5.2.3 Is Feedback Too Context-specific for Practical Reuse?

Schön persuasively argues that effective feedback should be context-specific and expert-generated [179]. He offers a vignette from architecture where the teacher suggests an alternative building to the student as an example of situated wisdom and its transfer. If Schön is right that this exchange requires both wisdom and context, does that mean that feedback reuse is infeasible? Within a given setting, project, or genre, common issues recur. Hewing to the principle of recognition over recall, we hypothesize that suggestions and guidance can increase novices' participation in context-specific exchanges.

5.2.4 Prior Systems & Approaches for Scaling Feedback

Existing approaches for scaling personalized feedback include clustering by similarity (*e.g.*, for writing [24] and programming [72, 90]). Gradescope [2] and Turnitin [3] allow graders to create reusable rubric items and comments to address common issues and apply them across multiple assignments. Gradescope binds rubric items to scores, which emphasizes grades rather than improvement.

Other methods include automating the reuse of the solutions of previous learners. These methods work best when correct and incorrect solutions are clearly distinct, such as in programming [71, 86] and logical deductions [57]. Automated methods have also found success with the formal aspects of more open-ended domains such as writing [24, 173]. However, assessing the quality and effectiveness of creative work – the strength of a design, the power of a poem – is intrinsically abstract and subjective and lies beyond current automated analysis techniques. Also, little automated analysis exists for media other than text. For domains like design, human-in-the-loop analysis will remain important for quite some time.

5.2.5   Automatically Detecting Feedback Characteristics

Although feedback is often specific and contextual [179], general characteristics can be automatically detected and used to help reviewers improve their feedback. For example, PeerStudio [129] detects when comments can be improved based on the length of the comment and the number of relevant words. Data mining and natural-language processing techniques can also automatically detect whether a comment is actionable or not, and prompt the reviewer to include a solution [154, 228]. Krause *et al.* use a natural-language processing model to detect linguistic characteristics of feedback and suggest examples to reviewers to help them improve their comment [126]. These methods require a reviewer to first submit their comment so it can be analyzed, and then improve their comment after submission.

5.3   CritiqueKit: Interactively Guiding Feedback

Based on these methods and insights, CritiqueKit (Figure 5.1) categorizes feedback and provides prompts and suggestions to reviewers. It differs from prior work by providing feedback to reviewers as they type rather than after they submit. We hypothesize that this contextual assistance may provide a just-in-time scaffold that changes how reviewers' thoughts crystallize, yielding feedback that is more specific, actionable, and/or justified.

Much like the previous systems in this dissertation, CritiqueKit's interface embeds assistance in the context of the user's task, by displaying interactive guidance and suggestions in a panel next to the work being reviewed. Like DiscoverySpace, CritiqueKit prioritizes ease of use for achieving outcomes; users can reuse a feedback suggestion by simply clicking on it. However, unlike DiscoverySpace, CritiqueKit also helps users improve the feedback they come up with from scratch. As our deployments will show, reviewers are often reluctant to directly reuse feedback suggestions, preferring instead to write their own ideas. While DiscoverySpace does not provide support for users who choose to use Photoshop's tools directly, CritiqueKit

explores how simple guidance might help reviewers go beyond direct reuse while still achieving outcomes quickly and easily.

5.3.1 Interactive Guidance as a Form of Scaffolding

CritiqueKit features an interactive guidance panel (Figure 5.1b) with checkboxes that update based on which of three attribute categories the feedback fits: *Is Specific* [126, 174, 194, 232], *Is Actionable* [50, 68, 129, 142, 174, 204], and *Is Justified* [70, 126, 151, 232].

The prototype assesses the feedback's fit with the following heuristics. The heuristic for the *specific* category merely requires that comments be at least five words long because vague comments tend to be short, such as "good job" or "needs work." Perhaps surprisingly, we observed that the five-word nudge was sufficient to garner specific feedback in practice. (Some websites, like Etsy, also use a five-word minimum heuristic for reviews). For the *actionable* and
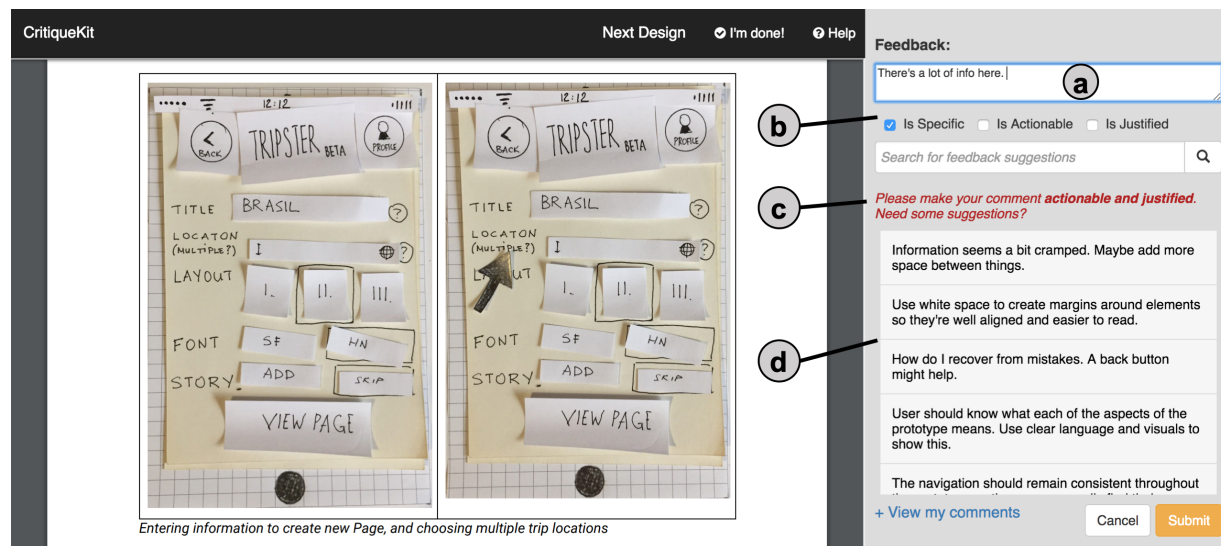
Figure 5.1: The final CritiqueKit interface (used for Experiment 2). a) The reviewer can type their feedback in the textbox. b) The checkboxes in the guidance panel update based on the characteristics of the reviewer's comments. c) CritiqueKit explicitly prompts reviewers to ensure their comment fits the checkboxes in the guidance panel. d) The reusable feedback suggestions in the suggestions box update based on the unchecked characteristics in the guidance panel, adapting specifically to the reviewer's feedback.

justified categories, we manually combed feedback that had been hand-labeled as meeting these categories and observed that specific keywords (i.e., "maybe try" and "you should" for actionable; "because" and "so that" for justified) were strong cues of these categories. Consequently, the prototype implementation simply checks for the presence of these keywords and phrases in feedback comments.

A comment is considered complete once all checkboxes are checked. Reviewers can manually check and uncheck the checkboxes if they feel the checkboxes did or did not add a category in error. For example, if a reviewer's comment states, "Use a 2-column grid layout," and the "Is Actionable" checkbox remains unchecked, the reviewer can manually check the checkbox to note that their comment does indeed contain an actionable suggestion.

5.3.2   Adaptive Suggestions for Greater Specificity

The suggestions box (Figure 5.1d) contains a list of previously given feedback from experts. These suggestions dynamically adapt based on how the reviewer's feedback is categorized in the guidance panel. For example, if a reviewer's comment does not yet satisfy the actionable and justified categories (as in Figure 5.1), the suggestions box would contain examples of feedback with these characteristics. Suggestions appear in the order they were added to the corpus.

5.3.3   The CritiqueKit Review Workflow

When a reviewer first opens CritiqueKit, a prompt asks them to provide specific feedback on something they like about the design and something that could be improved. The suggestions box contains general feedback snippets [128] pertinent to the review criteria to give reviewers a starting point, providing suggestions that are broadly applicable and fit within the specified criteria. The "Submit" button at the bottom of the interface is red to indicate that the comment text box is either empty or does not fit any of the categories in the guidance panel.

Once a comment is sufficiently long, the "Is Specific" checkbox will check, and the reviewer will be prompted to make their comment actionable and justified. The "Submit" button turns yellow to indicate that their feedback is not yet complete, though they can still submit if desired. The feedback suggestions then change to present comments that instantiate both actionable and justified feedback. The suggestions continue to adapt depending on the characteristics of the comment, showing reusable examples of feedback that satisfy the unchecked categories in the guidance panel. The reviewer can insert a suggestion directly into their own feedback by clicking on it. They can then edit the suggestion if they wish in the text field. Once all checkboxes are checked, the "Submit" button turns green as an indication of completeness.

Using prior feedback as suggestions can give inspiration and highlight common issues. The presence of the structured guidance panel reminds reviewers of attributes that feedback should have.

5.3.4   Implementation

CritiqueKit is a client-server web application implemented using Node.js; it assumes that all content to be reviewed is available on the web. The corpus of reusable feedback comments is stored on the server in JSON format.

CritiqueKit uses web sockets for communication between each client running the application and the main server, implemented using the socket.io module. Feedback classification happens on the client-side using JavaScript. Feedback suggestions are also generated on the client-side after retrieving the corpus from the server; the suggestions box adaptively shows and hides comments using JavaScript.

Users access CritiqueKit by navigating to its URL in a web browser. The first time the browser loads the website, a unique ID is generated for the user and sent to the server. A cookie is also saved on the client-side so that the server can identify and differentiate users. The review content is loaded within the page as an iframe.

5.4   Deployments: (How) Is Feedback Reused?

To understand how feedback is reused in educational settings and evaluate the CritiqueKit approach, we conducted two deployments and two experiments. All studies took place at a research university.

5.4.1   Deployment 1: How Do Teaching Assistants Reuse Feedback?

Eight teaching assistants (TAs) (two female) for an undergraduate design course used Gradescope to grade and critique seven weekly assignments that varied in content from storyboards to written explanations to high-fidelity web application prototypes. TAs set rubric items for each assignment and wrote comments for each. We deployed CritiqueKit to first understand how TAs might reuse feedback and made iterative improvements to the design throughout the quarter based on TA input.

Method: Integrating CritiqueKit with Gradescope

To integrate with the TAs' existing workflow, we implemented CritiqueKit as a Google Chrome extension that augments the Gradescope interface with a suggestions box (Figure 5.2). This version of CritiqueKit contained only the suggestions box to explore feedback reuse. The suggestions box contained a manually curated set of feedback provided by former TAs in a previous iteration of the course, stored in a Google Sheet online and retrieved by the Chrome extension using the Google Sheets API. Suggestions were categorized into three feedback categories: Positive, Problem, and Solution. TAs could select feedback suggestions to directly copy into the textbox for further editing. Each rubric item contained its own suggestion box interface, providing suggestions specific to that rubric item.

We curated the reusable suggestions corpus as follows. Given all feedback from the previous quarter, feedback that was 25 or fewer words in length was kept, because longer

feedback was both too long to be skimmed in a suggestion display and tended to be overly specific. Feedback of 26-30 words was truncated at the sentence level to fit within the 25-word limit. Longer comments or duplicate comments were discarded. In total, 526 comments were provided as suggestions throughout the course for seven (of ten) assignments. Suggestions were manually categorized into the Positive ($n = 92$), Problem ($n = 312$), and Solution ($n = 122$) categories.

Result: TAs Used Feedback Suggestions as Inspiration

Across seven assignments, four of the eight TAs reused 51 distinct suggestions from the 526-element corpus (9.7%). 75 of 583 designs received a reused suggestion for feedback. 60%
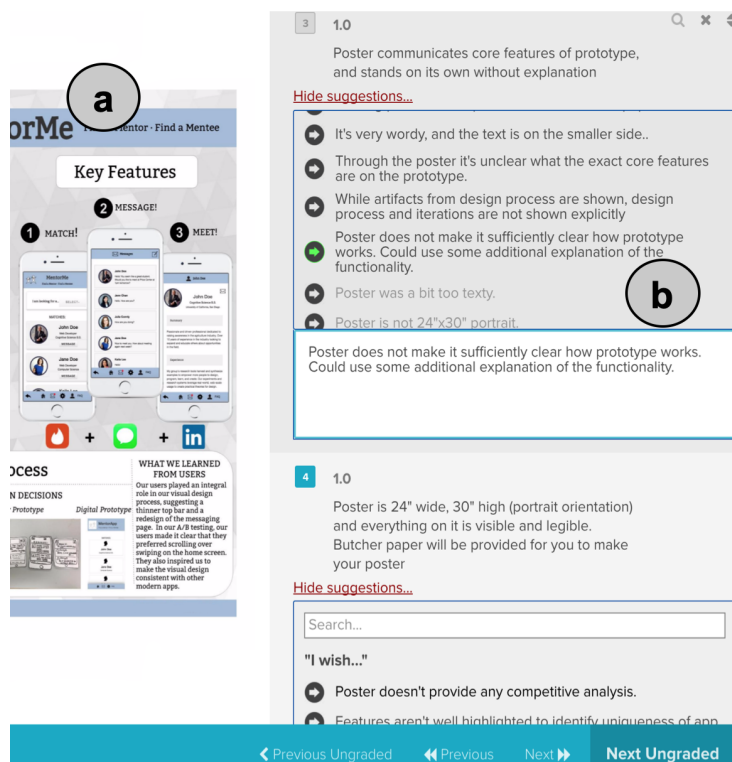
Figure 5.2: CritiqueKit implemented as a browser extension in Gradescope for Deployment 1. a) Reviewers provide feedback on a student design. b) The suggestions box under each rubric item provides reviewers with a list of reusable suggestions and a comment box for providing feedback on a submission.

of reused suggestions were categorized in the Problem category. These numbers omit any reuse occurring entirely inside Gradescope without CritiqueKit. (Gradescope provides an interface for reusing entire comments within an assignment rather than for individual parts of the comment.)

An end-of-course survey asked TAs about their CritiqueKit use. One commented that he would "*skim the comments in the [suggestions] to see if something was accurate to my thoughts*." Another mentioned that the prototype helped him "*[find] ways to better explain and give feedback about specific points*." TAs also mentioned that suggestions sometimes reminded them to comment on more diverse aspects of students' work. For example, one mentioned that seeing positive suggestions reminded her to give positive feedback, not only critiquing areas for improvement. TAs mentioned using the suggestions as inspiration rather than the exact wording, taking the underlying concept of a suggestion and tailoring it.

5.4.2   Deployment 2: How Do Students Reuse Feedback?

The first deployment examined teaching staff usage; this second deployment examined student usage to understand how novices interact with guidance and suggestions. We deployed CritiqueKit as a standalone web application with 29 students in an undergraduate design course for five weeks. Students gave anonymous feedback on two randomly assigned peer submissions for each of seven assignments.

Method: Integrating Interactive Guidance for Scaffolding

Novice students are less experienced in giving feedback and may benefit from interactive scaffolding [168]. This version of CritiqueKit included an interactive guidance panel to help reviewers provide more specific and actionable feedback (Figure 5.3). The categories on the guidance panel were "Is Positive," "Is Specific," "Identifies a Problem," and "Presents a Solution" with checkboxes next to each. These categories stem from recommendations in the literature for both positive and critical feedback [111]. Similar to the final version of CritiqueKit, these

checkboxes updated as a reviewer typed by classifying their comment into the three categories. The categories differed from the final version, focusing on specific and actionable feedback.

The suggestions box was seeded with feedback from the course TA. Similar to the first deployment, the suggestions were categorized in the Positive, Problem, and Solution categories. When a student submitted a comment, it was classified into one of these categories, shortened to 25 words if it was longer, and fed back into the corpus to appear as a suggestion, enabling students to reuse their peers' as well as their own comments. The suggestions were ordered first by frequency used, then by shortest length first, and updated as these values changed and more comments were added. Compared to the final version of CritiqueKit, suggestions were static, meaning they did not change as the reviewer typed.

Results: Positive Feedback Common; Reuse Rare

For seven assignments, 898 comments were submitted. Independent raters classified each comment into the five categories of Positive Only, Positive and Specific (Positive + Specific),
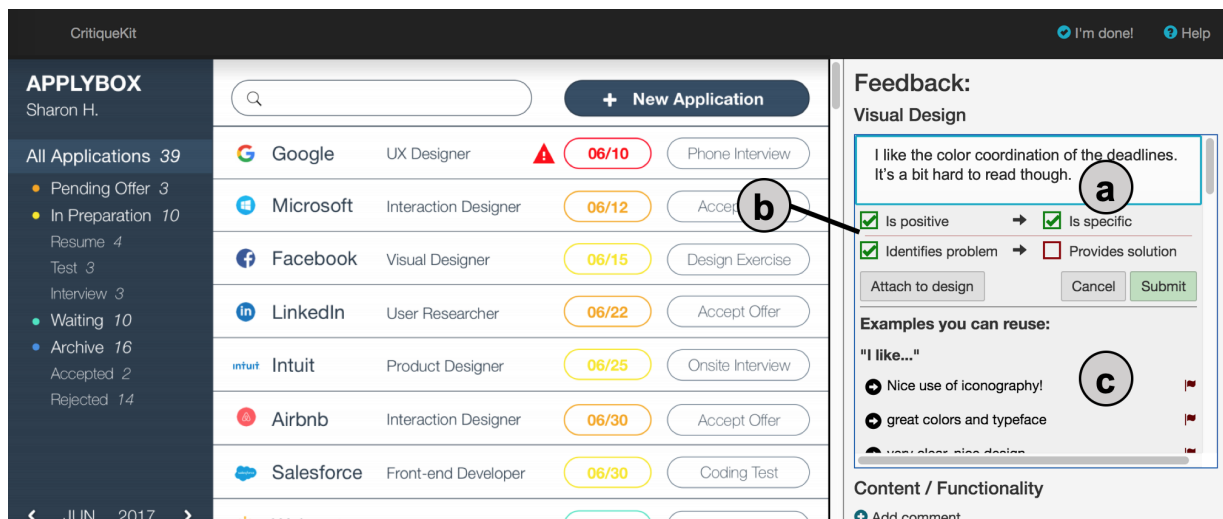
Figure 5.3: The CritiqueKit user interface for Deployment 2 and Experiment 1. a) The reviewer types their feedback into the text box. b) Checkboxes in the guidance panel update as the reviewer types to show how well the comment fulfills high-quality feedback criteria. c) The reviewer can browse and reuse previously given feedback.

Problem Only, Solution Only, and Problem with a Solution (Problem + Solution). 45% of these comments contained positive feedback; 30% contained a Problem + Solution statement.

Students rarely selected feedback suggestions for reuse. Over the five-week deployment, 14 distinct suggestions were reused on 27 student designs for four of the seven assignments. These suggestions were mostly short, vague comments such as "I wish this was more visually appealing." This may be because students often left feedback specific to individual designs that did not easily generalize to other contexts. Students' comments in a post-survey confirmed that the suggestions did not always seem applicable. Students also did not regularly use the interactive guidance panel; 15 of the 29 students engaged with the panel a total of 120 times over five weeks.

In contrast to how TAs reused feedback, students may not have recognized common issues. TAs paid attention to common errors between designs and mainly reused Problem feedback, whereas students may not have noticed or attended to underlying issues between designs. For instance, one student mentioned that they did not use the feedback suggestions because they "*rarely pointed out the same things when critiquing interfaces*."

This exploratory deployment investigated how students reuse feedback and respond to interactive guidance in the classroom. To understand how a system with these features compares to a standard feedback system, the next study was a controlled between-subjects experiment.

5.5   Experiments: Scaffolding Feedback

Following our deployments, we conducted two empirical studies to investigate the impact of suggestions and guidance on feedback quality.

5.5.1   Experiment 1: Do Static Suggestions Improve Feedback?

In an online between-subjects study, 40 undergraduate design students were asked to review three restaurant website homepages using CritiqueKit. The task emulated peer review

tasks often required in creative courses. This study's suggestion corpus came from a design feedback task on CrowdCrit [142] and was categorized in the Positive, Problem, and Solution categories. We hypothesized that suggestions and guidance would help reviewers provide more specific and actionable comments.

Method: Reviewing Restaurant Websites

40 participants were randomly assigned to either the CritiqueKit condition or the Control condition (20 in each). CritiqueKit participants used the same version of CritiqueKit as Deployment 2 with all features available (Figure 5.3). Control participants used an otherwise identical version consisting solely of a comment text box. Upon landing on the homepage of either version, participants were provided with a scenario explaining that three restaurant owners are seeking feedback on their new website design. Participants were given a brief tutorial of CritiqueKit's features and an explanation of what makes for good feedback. There were no restrictions or requirements on time spent or amount of feedback to provide. We compared the percentages of comments in five categories. Comments including a supportive element were labeled as *Positive Only* or *Positive + Specific*. Comments including a critical element were labeled *Problem Only*, *Solution Only*, or *Problem + Solution*.

Results: Static Suggestions Were Not Helpful

With static suggestions and interactive guidance, there were no significant differences between conditions. (To foreshadow, we will see differences in Experiment 2, which adds adaptive suggestions). Participants provided a total of 323 comments (168 for control, 155 for CritiqueKit). The average number of words per comment was not significantly different between conditions (Control: $m = 29.07, SD = 23.64$; CritiqueKit: $m = 23.22, SD = 17.3$) ($F(1,38) = 2.52, p = .11$).

Suggestions & Guidance Did Not Affect Type of Feedback

The distribution of the five category types did not vary significantly between conditions ($x^2 = 4.80, df = 4, p = .31$) (Figure 5.4). In both groups, participants provided mostly Problem + Solution feedback (39% in Control; 34% in CritiqueKit).

Most CritiqueKit Participants Corrected Category Labels

65% of CritiqueKit participants actively used the guidance panel, making a total of 85 corrections to categories. Interaction with the guidance panel may have indicated attention to the feedback characteristics. As the study was online, we don't know how many of the other 35% were influenced by the guidance panel.
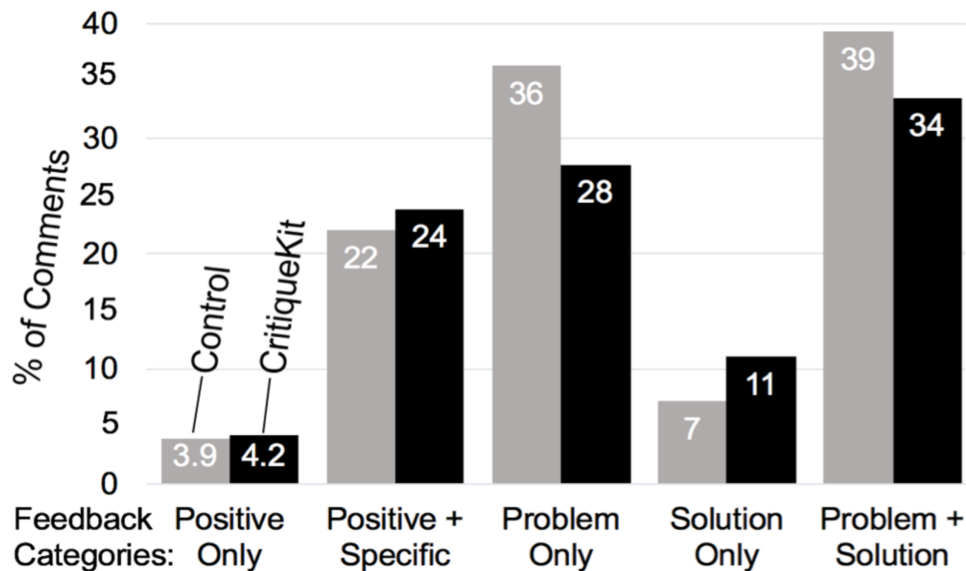
Figure 5.4: A plurality of feedback in both conditions in Experiment 1 identified both a problem and solution (*i.e.*, was actionable). Feedback that was only positive was the rarest. There were no significant differences between conditions for these categories.

Unfortunately, People also Reused Vague Suggestions

11 distinct suggestions from the corpus were reused. 8 of these were vague; 3 were specific. 15 of 155 reviews included a reused suggestion. This seems especially low given the high engagement with the guidance panel. We see two reasons for this: First, the suggestions came from CrowdCrit [142], where participants provided feedback on a weather app design. The study task was different than the task for which the suggested feedback was originally given, and novices may have had a limited ability to see the deep structure behind a suggestion and reapply it in a new context. Second, the suggestions were created by crowd workers and of uneven quality.

The suggestions selected were typically short, positive comments, perhaps because students did not know how to apply them in the specific context. For example, the most commonly reused suggestion was "great use of color" (reused 3 times). This result is similar to Deployment 2 in which students did not find feedback provided by other peers or novices to be useful and generalizable. Feedback suggestions may require more curation or quality control to be most useful.

Suggestions & Guidance Should Work in Concert

While this version of CritiqueKit contained both feedback suggestions and interactive guidance, these features functioned independently. Regardless of the categories checked in the guidance panel, the suggestions remained static and presented in the same order for each participant, potentially making them easy to ignore if they were irrelevant to the context. Participants may have paid attention to only one feature at a time. The next study investigated the question of whether adaptively presenting feedback suggestions along with interactive guidance improves feedback.

5.5.2 Experiment 2: Do Adaptive Suggestions Help?

The second experiment used the final version of CritiqueKit described in the system section to test the hypothesis that adaptively-presented suggestions combined with guidance would improve feedback by increasing the fraction of feedback that is specific, actionable, and/or justified.

Method: Reviewing Paper Prototypes

We conducted a between-subjects in-person study with 47 (27 female) participants. Participants were recruited from an undergraduate subject pool within the Psychology and Cognitive Science departments. Participants were randomly assigned to either the CritiqueKit ($n = 24$) or Control ($n = 23$) conditions. 44 of these participants had no design course experience; 3 participants had taken at least one design course. 28 spoke English as a second language.

Participants were asked to provide feedback on two designs from students enrolled in an online course who volunteered to receive more feedback on their work. These designs were PDFs of mobile application paper prototypes. The review criteria included whether the prototype supported the student's point of view and whether it seemed easily navigable. Participants were first shown the design instructions and review criteria and then given a short tutorial of CritiqueKit as well as an explanation of what makes for good feedback. CritiqueKit participants had all features of CritiqueKit available to them (Figure 5.1), while Control participants used a version that consisted of only a textbox for their feedback. The task took about 30 minutes to complete. After providing feedback on both designs, participants were interviewed about their feedback process and use of CritiqueKit.

Presenting Feedback Suggestions Adaptively

The categories on the guidance panel and their definition used for coding participants' responses were the following:

Specific: relates directly to the review criteria

Actionable: gives a concrete suggestion for improvement

Justified: provides a reason for why something is done well or should be improved

For Deployment 2 and Experiment 1, the guidance panel categories sought to encourage specific and actionable feedback (Figure 5.3). Examining the feedback from our previous studies, we found that "Is Positive" and "Identifies a Problem" did not provide significant guidance as reviewers were generally aware of whether their feedback was positive or critical. In addition, the guidance panel did not explicitly check for justification of feedback. For Experiment 2, we revised the categories to "Is Specific," "Is Actionable," and "Is Justified" to also encourage the explanation or reasoning behind feedback. As described in the system section, the checkboxes update as the reviewer types to reflect the categories present in their comment, and the suggestions adapt to show feedback examples from categories not yet present in the comment.
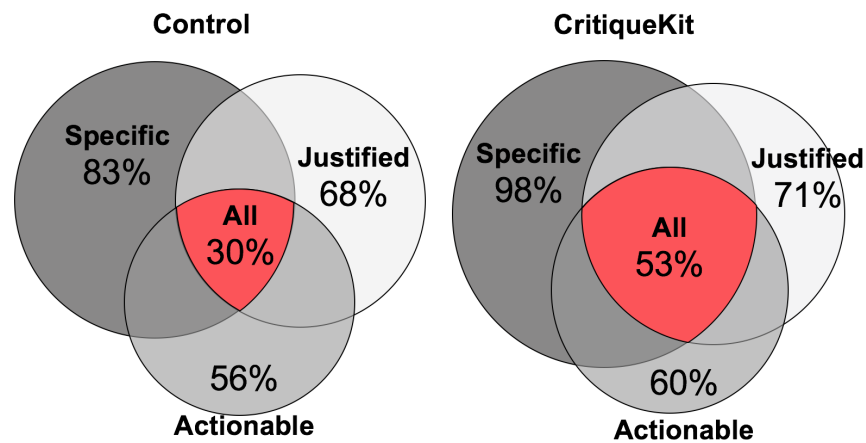
Figure 5.5: In Experiment 2, a significantly higher percentage of feedback in the CritiqueKit condition ($n = 24$) (53% versus 30%) contained three attributes of good feedback: Specific, Actionable, and Justified ($x^2 = 8.33, df = 1, p = .01$) than in the Control condition ($n = 24$).

Results: CritiqueKit Participants Provided a Greater Proportion of Specific, Actionable, & Justified Feedback

Participants provided 158 total comments (79 control, 79 CritiqueKit). The percentage of comments that contained all three categories (specific, actionable, and justified) was significantly higher in the CritiqueKit condition (53%) than in Control (30%) ($x^2 = 8.33, df = 1, p = .01$) (Figure 5.5). As an example, this comment meets all three: "The 'more questionnaires' section (Specific) should be made smaller (Actionable) because it is not the main focus of the page." (Justified). The system's heuristic for checking specificity of a comment was quite simple: five words or greater in length. Feedback raters blind to each condition used a more sophisticated and holistic assessment, taking specific to also mean related to the review criteria. With this assessment, 98% of CritiqueKit comments were labeled by raters as specific whereas only 83% of Control comments were. These raters also rated comments from Experiment 1 within the specific, actionable, and justified categories to provide a comparison between the two experiments. Interestingly, the percentage of comments containing all attributes in the Control condition was relatively consistent between Experiment 1 (35%) and Experiment 2 (30%). The percentage of comments with all attributes in the CritiqueKit condition greatly increased between the two experiments (26% versus 53%). Having the checkboxes may have explicitly reminded CritiqueKit participants to ensure their comments satisfy the specific, actionable, and justified categories.

Because longer comments were more likely to contain all three categories, each comment was also scored on a point scale and averaged per participant. Comments received one point for each specific, actionable, and justified idea (Figure 5.6). A MANOVA with category points as dependent variables shows a significant difference between conditions ($F(1,3) = 3.21, p < .005$). CritiqueKit participants provided more specific ideas than Control participants (Control $m = 3.87$, CritiqueKit $m = 5.17$, $F(1,156) = 14.04, p < .05$). This may be because the suggestions provided examples of relevant ideas and led CritiqueKit participants to address more. CritiqueKit participants also provided more ideas that fit all three categories than control (Control $m = 1.0$,

CritiqueKit $m = 2.2$, $F(1, 156) = 8.78$, $p < .005$). Given that most participants did not have any design experience, the combination of adaptive suggestions and guidance may have been most useful for these reviewers. The suggestions may have provided a starting point while the guidance panel helped them understand how to apply the attributes of good feedback. There were no significant differences in the average number of actionable and justified ideas in comments.

On average, Control comments were 39.3 ($SD = 30.3$) words long and 43.7 ($SD = 31.4$) words for CritiqueKit comments. There was no significant difference in comment length ($F(1, 156) = 1.77$, $p = .19$). Unfortunately, we don't know how the feedback improved students' work because they received feedback from both Control and CritiqueKit participants. A longitudinal deployment with the final version of CritiqueKit would likely be more useful in determining the helpfulness of feedback.

Suggestions Helped Reviewers Describe Their Thoughts

Participants rated the suggestions as being generally helpful ($m = 4.29$, $SD = 0.95$, 1-5 Likert Scale). When asked to elaborate on their rating, many participants noted that the suggestions helped them describe their thoughts. One participant remarked, "*I was a bit lost*
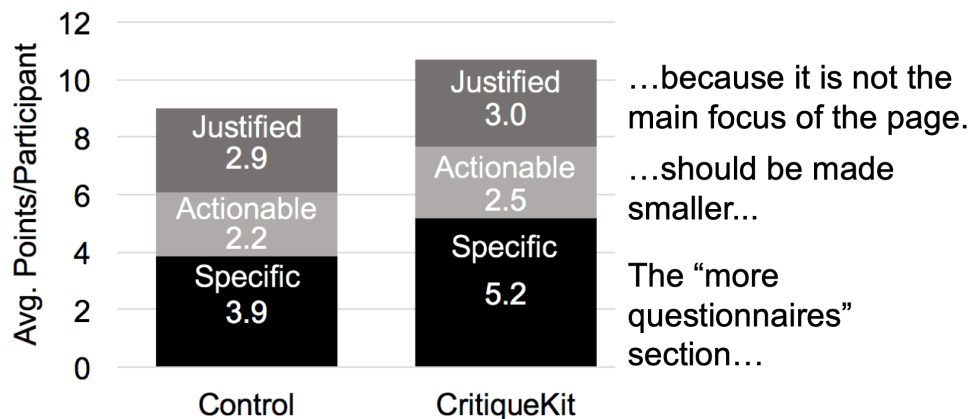
Figure 5.6: CritiqueKit participants ($n = 24$) provided more *specific* and *all-three* category ideas than Control participants ($n = 23$)($F(1, 156) = 8.78$, $p < .005$).

at first because I didn't know how to describe my thoughts. The suggestions helped me figure out how I should describe what I was thinking." Similarly, another mentioned that "*when I [didn't] know how to put my feedback in words, I could look at the suggestions.*" Particularly for participants without any design experience, suggestions helped with appropriate language to use in their feedback. For example, one noted that "*seeing actual wording from a designer's point of view was good so you know how to say what you want to say.*" Though a few participants did not directly select suggestions, it is likely that they were inspired or influenced by them as they used similar wording in their own comments.

Still, some participants felt the suggestions were too general and not entirely relevant to the specific design they were reviewing. One participant felt constrained by the suggestions, stating that she ignored them because she wanted to write her own opinions. Suggestions seemed most helpful for participants who used them as a starting point for their own thoughts rather than solely relying on them. Participants who simply selected suggestions tended to list issues without adding their own elaboration. This behavior not only led to incomplete feedback, but also produced depersonalized and scattered comments. For example, one comment that solely relied on suggestions reads "User immediately knows the purpose of the prototype. Good use of grid layout to keep items aligned. Icons should be immediately recognizable to the user." A consideration for future work is to develop feedback suggestions tailored to help reviewers provide more cohesive and contextual comments.

Most participants in this experiment did not have any design experience and may have benefited most from the suggestions. Many participants noted using the suggestions as a way to find ideas whereas students with design experience may already have heuristics and processes in mind when providing feedback. Future work should examine how suggestions and guidance might improve feedback for more experienced learners as well.

Interactive Guidance Helped Remind & Focus Reviewers

Participants were mixed on the helpfulness of the guidance panel ($m = 3.67, SD = 1.2$,

1-5 Likert Scale). Those who did find it helpful noted that the categories helped guide their

feedback process. For instance, one participant noted that he "*went in order of the checkboxes.*

First, I provided something specific, then something actionable, then justified it." Another noted

that the categories helped her know whether her feedback was actually useful or helpful, and one

noted that the guidance panel "*[made] sure the feedback is complete and not vague.*"

Anecdotally, we observed that when participants said the categories were not useful, it

was because they believed them to be inaccurate in their classifications. The accuracy (compared

to human raters) for the actionable category was 67% and 75% for the justified category. A

participant stated that "*[the checkboxes] didn't always check when I thought they should, so I*

would just do it myself." Another thought the checkboxes were "*quick to judge, it felt like it*

wasn't reading what I was saying." Three participants, who were not native English speakers,

found the categories confusing because they weren't sure what they meant. Future iterations of

CritiqueKit could include the definition of these categories in the prompts to make the meaning

clearer. Interestingly, a couple participants noted that they used the categories as reminders

rather than for active guidance. For instance, a participant mentioned that though he felt the

interactive guidance was not that accurate, "*[the checkboxes] reminded me to make sure my*

comment contained specific, actionable, and justified parts, so I'd go and reread through my

comment."

Some participants commented on the adaptive presentation of the suggestions with the

guidance panel. For one participant, the suggestions helped him understand what the categories

meant. He noted, "*The whole actionable and justified thing, I didn't know what that meant, so*

the suggestions helped with that." Observations of participants showed that some clicked on the

checkboxes simply to see the suggestions under each one. When asked about how useful they

found CritiqueKit in general, participants varied widely in their ratings of usefulness. A more

precise measure would allow participants to compare across conditions, which was not possible with this between-subjects design.

5.6   Discussion & Future Work

This chapter empirically investigated two techniques for scaffolding feedback: reusable feedback suggestions and adaptive guidance. This work complements this dissertation more broadly, highlighting the benefits of adaptive guidance for interfaces that support creative tasks. Here we discuss and synthesize the findings.

5.6.1   Generating Reusable Feedback Suggestions

This work investigated whether suggestions and guidance can scaffold the feedback process. For this strategy to work, an eye towards reuse and adaptive feedback must be adopted. As Schön argues, experts may be most capable of recognizing common patterns and giving useful feedback [179]. However, while feedback should be specific, underlying concepts can generalize across contexts. In the studies that used expert-generated feedback suggestions (Deployment 1 and Experiment 2), participants cited the same reason for why the suggestions were useful: as inspiration. Participants reported that the suggestions helped them find words for their thoughts or helped direct their attention to issues they did not originally notice. This suggests that reusable suggestions should focus attention to common issues rather than specific instances. Our approach demonstrates how expertise on creative work can be scaled by providing feedback on a few to apply to many [128]. This extends work on reusable feedback in coding and writing [24, 86, 90] while keeping the human in the loop, enabling novices to learn and reuse expert insights.

It is possible that more general suggestions can lead to less personalized feedback, particularly in abstract domains like visual design. We observed this in 7 of the 79 comments from the CritiqueKit condition in Experiment 2, in which the four participants simply selected

92

suggestions without further elaboration. A consideration for creating and presenting reusable suggestions is how these suggestions can be both general yet personal to be more helpful to the recipient.

5.6.2 What is the Best Way to Guide Feedback?

Prior empirical work on feedback (*e.g.*, Kulkarni *et al.* [128] and Krause *et al.* [126]) has not compared static and adaptive suggestions. In this chapter, we found that people rarely used static suggestions and did not find them helpful; adaptive suggestions were used more and found more helpful. This reinforces prior work demonstrating that adaptive presentation of examples can improve learning [134, 150]. By presenting feedback suggestions that directly addressed missing characteristics of a reviewer's feedback, reviewers were prompted on where they could specifically improve, and explicitly shown examples of how to do so.

The second experiment adapted feedback suggestions based on whether their feedback was categorized as specific, actionable, and/or justified. Though some of the prototype's categorizations were misleading or inaccurate (for example, the comment "user flow is simple" was categorized as "Is Actionable" because of the word "use", even though it lacks a concrete suggestion), participants still referenced the three categories when composing their comments. The guidance panel was useful as a reminder to include the attributes of good feedback in their comments. A more sophisticated method for categorization would likely be helpful, though our naïve approach performed reasonably well overall.

The guidance panel focused on three important attributes of good feedback. A consideration is to also provide guidance for emotional content in feedback, as emotional regulation is important to how learners perceive feedback [126, 214]. In addition, other characteristics may also contribute to perceived helpfulness, such as complexity or novelty [126], that could be further explored through adaptive guidance.

5.6.3 Creating Adaptive Feedback Interfaces

In order for adaptive guidance to be most effective, the interface should be suitable for adaptation. In the two deployments and first experiment, the suggestions were not curated in any way: more than 1,400 comments were supplied as suggestions, but only 76 of these were reused by reviewers. Having more suggestions available was not beneficial because the suggestions were not sufficiently adaptable and were potentially irrelevant and difficult to browse. Experiment 2 introduced a curated approach: experts provided the suggestions with generalizability in mind. Of the 47 suggestions created, 29 were reused. Though fewer suggestions were available, they were more general and adaptable, potentially making them more useful.

Suggestion presentation shares many properties with search interfaces. Like with search, a good result needs to not only be in the set, but toward the top of the set [91]. The second experiment contained fewer suggestions, enabling easier search and browsing. Effective curation and display of suggestions should take into consideration the quality of feedback suggestions and how likely they are to be selected, potentially using frequency or some measure of generalizability as a signal.

Summary

Looking across the deployments and experiments, adaptive suggestions and interactive guidance significantly improved feedback while static suggestions did not offer significant improvements. These techniques were embodied in the CritiqueKit system, used by 95 feedback providers and 336 recipients. Although CritiqueKit's main goal was to help reviewers produce better outcomes quickly and in the moment, we suspect that the combination of interactive guidance and illustrative examples may also help reviewers learn and retain procedural about writing good feedback. Future work should evaluate this hypothesis and further investigate how best to create, curate, and display helpful suggestions.

Much knowledge work features both underlying principles and context-specific knowledge

of when and how to apply these principles. Potentially applicable feedback and review areas include domains as disparate as hiring and employee reviews, code reviews, product reviews, and reviews of academic papers, screenplays, business plans, and any other domain that blends context-specific creative choices with common genre structures. More generally, this work showed how contextual suggestions and guidance can help improve creative outcomes in the moment by leveraging and reusing existing expert work. We hope that creativity support systems of all stripes will find value in the ideas and results presented here.

Acknowledgements

Chapter 6

Future Explorations in Creative Problem-Solving

This dissertation presented three strategies for helping people accomplish their creative goals from the stages of getting started to giving and receiving feedback for improvement. This chapter presents preliminary experiments in future directions for catalyzing creativity and using exploration as a learning strategy for children and design students.

6.1   Exploratory Thinking in Children through Patterning

Children are extraordinary explorers. Unprohibited by set structures and knowledge of available representations [96, 206, 225], children can come up with more unusual hypotheses and explanations than adults [77, 141, 226]. Interestingly, where adults tend to fixate and satisfice [101, 113, 157, 190], children demonstrate cognitive flexibility and even a bias towards exploration [77, 141, 196, 197]. Given children's propensity for exploration, how might exploration be used as a pedagogical tool? Compared to direct instruction or worked examples, exploration strategies can support better engagement and discovery learning, leading to improved transfer of knowledge to new problems [23, 73, 185]. However, pure exploration and unguided learning can

be inefficient and induce a heavy cognitive load [121, 205]. We hypothesize that for the domain of pattern learning, directed exploration will yield more productive results [109, 226].

We investigate exploration as a learning strategy compared to a demonstration of a pattern for preschool and kindergarten-aged children. Patterning is an important skill for children, particularly in learning algebraic mathematics [161]. Young children may gravitate toward patterning activities because no alphanumeric knowledge is needed to engage in them. Prior studies show that children as young as four years old can recognize and abstract (create the same pattern using new materials) repeating pattern units [63, 171]. Though patterning interventions improve children's mastery of specific patterns, they still often struggle to transfer their knowledge to new contexts, especially when assessed on their ability to recognize the unit of repeat [63, 160]. Given the success of exploratory learning strategies on knowledge transfer in other domains of learning [23, 73, 185], we compare whether exploration will facilitate implicit pattern recognition and transfer of patterning ability more than a demonstrated example of the pattern.

6.1.1 Experiment: Can Exploration Improve Pattern Learning?

Participants

We recruited participants via emails collected from an online research database. 68 children, aged five to six years old (mean age= 5.98, $SD = 0.58$) participated in this study. Of these children, 34 were five-year-olds, and 34 were six-year-olds (10 children were excluded from analyses due to failing pretest items or technical difficulties that prevented completion of the study). Children were split into either the *exploration* or *demonstration* conditions. In the *exploration* condition, participants engaged with the pattern materials directly while *demonstration* condition participants watched the experimenter engage with the pattern materials. The experiment design was yoked, meaning that for each participant in the *exploration* condition, a paired participant in the *demonstration* condition would see the same information in the same order as the participant

in the *exploration* condition. All pairs of participants were age-matched.

Procedure

The experiment took place over the videoconferencing platform Zoom with a prototype developed in Google Slides to allow for synchronous observation and communication between the child participant and experimenter. Parents were instructed to not interfere with the child's performance unless in the case of technical difficulty. The patterns within the prototype consisted of various shapes created with the Shapes tool in Google Slides. Children interacted with the shapes by clicking and dragging them, with help from the experimenter or parent if needed. After a pattern duplication task, the experiment consisted of a pattern abstraction pre-test and a pattern learning task. A pattern transfer and pattern abstraction post-task followed pattern learning. After the full experiment, children and parents were debriefed and received a $5 USD gift card for their time and participation in the 10-15 minute experiment. Pattern duplication and abstraction tasks were adapted from [172]; pattern learning and transfer tasks were novel tasks designed for this study.

Tasks

Pattern Duplication Task. Participants saw a model two-unit AAB pattern and were asked to duplicate the pattern using materials of the same shape and color. The pattern was considered correct if at least one unit of the solve pattern matched the model pattern. This task served as exclusionary criteria for the study.

Pattern Abstraction Pre- & Post-Tests. The abstraction pre- and post-tests consisted of a two-unit ABB model pattern and a two-unit AAB model pattern. The experimenter asked children to create the same pattern using materials of different shapes and colors than the model pattern. The pattern was considered correct if at least one unit of the solve pattern matched the model pattern. Children completed the abstraction tasks both before and after the pattern learning and
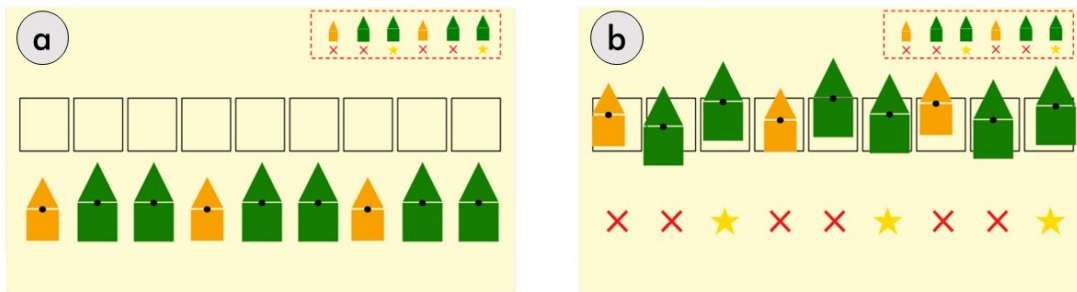
Figure 6.1: a) The pattern learning task consists of three units of an ABB pattern with house shapes and a small icon to remind children of the underlying pattern. b) Children can reveal the contents of the houses to to find the hidden stars underneath. *Exploration* condition children chose the order in which the houses were revealed while *demonstration* condition children watched the experimenter reveal the houses in the order of their age-matched *exploration* counterpart.

transfer tasks as a measure of learning or improvement.

Pattern Learning Task. The pattern learning task consisted of a row of nine house shapes of two sizes and colors: the orange houses were small shapes, and the green houses were big shapes. The nine houses formed a three-unit ABB pattern (small, large, large) (Figure 6.1a). The second 'B' shape in each ABB unit contained a star shape hidden underneath with all other house shapes hiding an 'x' underneath (Figure 6.1b). We informed all participants that their goal was to find the three stars by figuring out the pattern to find them. Participants in the *exploration* chose the order in which to reveal the hidden contents of the nine house shapes. For the *demonstration* condition participants, the experimenter moved all shapes for the participant in the same order as the participant's paired *exploration* condition counterpart.

Pattern Transfer Task. The pattern transfer task consisted of the same ABB pattern row as the pattern learning task, but with different shapes of different colors (red and purple lightbulbs)(Figure 6.2a). As with the learning pattern, stars were hidden under the second B shape in each ABB unit (Figure 6.2b). The experimenter explained that the participant could find the stars in the same place using the same pattern as the pattern learning task. A small thumbnail in the corner of the slide served as a reminder of the previous pattern from the learning task.
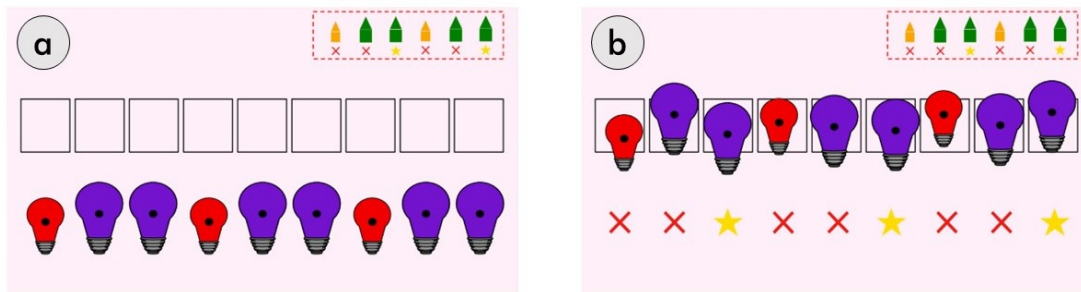
Figure 6.2: a) The pattern transfer task consisted of the same three-unit ABB pattern as the pattern learning task with different shapes. b) Children had five chances to reveal any of the shapes to find the three hidden stars. A small icon in the corner reminded children of the pattern from the pattern learning task.

The experimenter gave children five chances to try to find the three hidden stars, giving verbal feedback after each attempt. All children were told to choose five shapes even if they found all three stars prior to the fifth attempt.

6.1.2   Preliminary Results

No significant difference in pattern abstraction

Pre- and post-test tasks were scored as a binary outcome for whether the pattern was correct or incorrect. We combined the two pre-tests and the two post-test scores into one pre-test and one post-test score with a total out of two for each. Children in both conditions performed nearly at ceiling for both pre- and post-abstraction tasks (*exploration* pre-test *mean* = 1.79, *SD* = 0.54; post-test *mean* = 1.82, *SD* = 0.52. *demonstration* pre-test *mean* = 1.91, *SD* = 0.29; post-test *mean* = 1.94, *SD* = 0.24). A linear regression with age as a covariate shows that the average difference from pre- to post-test was not significantly different ($F(2,65) = 1.76, R^2 = 0.05, p = .18$).
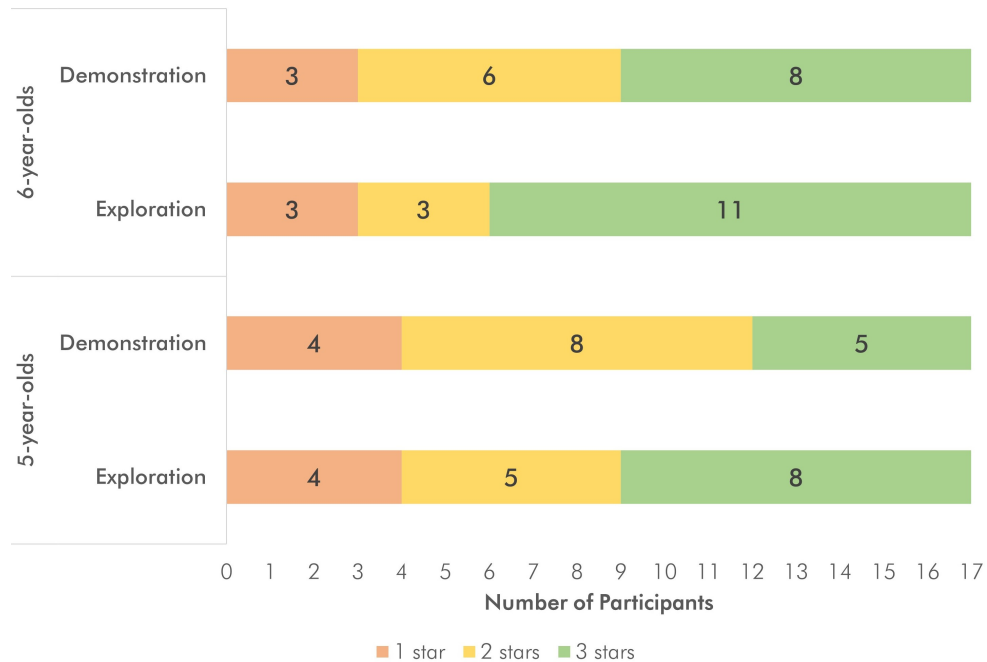
Figure 6.3: The number children who found one, two, and three hidden stars in the pattern transfer task ($n = 17$ for each age group in both conditions; $n = 34$ total in each condition). In both conditions, more six-year-olds successfully found all three stars compared to five-year-olds.

Six-year-olds recognize patterns better than five-year-olds

In general, six-year-olds performed better on the pattern transfer task, with a larger proportion finding all three stars, demonstrating an understanding of the unit of repeat in the ABB pattern. In the *exploration* condition, 8 of the 17 five-year-olds found all three stars compared to 11 of the 17 six-year-olds. In the *demonstration* condition, 5 of the 17 five-year-olds found all three stars compared to 8 of the 17 six-year-olds. All participants found at least one star in the task. Figure 6.3 shows the number of participants who found one, two, and three stars by age and condition. Though there are no significant differences between conditions in the preliminary data ($\chi^2 = 2.76, df = 2, p = .25$), we observed a slight trend towards better performance in the *exploration* condition.
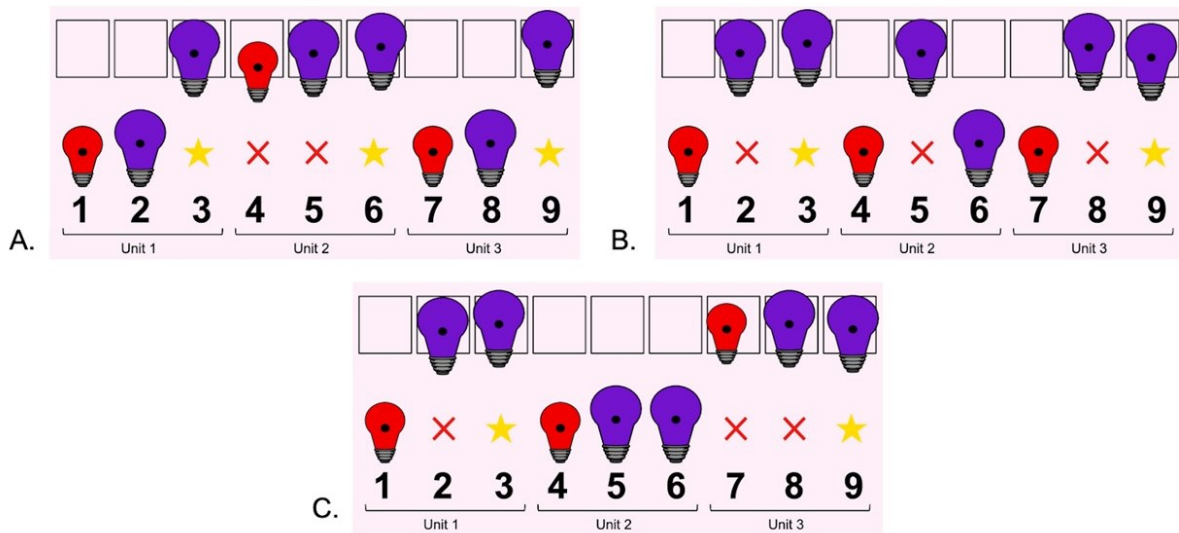
Figure 6.4: Examples of the pattern-relevant strategies observed in the pattern transfer task: A) Unconstrained chunking - searching under at least one shape from all three repeated units, B) Shape matching - searching only under the large shapes, and C) Constrained chunking - searching under shapes from only two of the three repeated units.

Exploration affects pattern learning strategies

Though there were no significant differences in performance on the pattern abstraction task, children interestingly used different strategies. One researcher and a research assistant coded children's patterning strategies blind to condition according to three main categories: pattern-irrelevant, pattern-relevant, and advanced pattern strategies. Disagreements and sub-strategies were discussed and resolved between the two raters. A sequential strategy (choosing shapes from either left to right or right to left in a sequential order) and spatial grouping (searching only in closely proximal shapes) represented the two sub-strategies in the pattern-irrelevant strategies as they do not demonstrate an understanding of the underlying pattern. Unconstrained chunking (searching at least one shape in each of the three repeating units)(Figure 6.4A), shape matching (searching only in the B shapes in each ABB unit)(Figure 6.4B), and constrained chunking (searching at least one shape in only two of the three repeating units)(Figure 6.4C) represented the three pattern-relevant strategies as they demonstrate some understanding of the underlying

Table 6.1: Children engaged in various strategies to solve the pattern transfer task. In both the *exploration* and *demonstration* conditions, six-year-olds engaged in more mature, pattern-relevant strategies than five-year-olds.

Strategy Type	Sub-Strategies	5-year-olds		6-year-olds	
		Explore	Demonstrate	Explore	Demonstrate
Pattern-Irrelevant	Sequential	1	5	2	1
	Spatial Grouping	3	2	0	1
Pattern-Relevant	Unconstrained Chunking	7	5	7	8
	Constrained Chunking	1	1	0	1
	Shape Matching	2	2	1	2
Advanced	Exploit	3	2	7	4

pattern. Exploit (finding all three stars in a row) represented the most advanced strategy as it demonstrates a clear understanding of not only the underlying pattern, but also the understanding of where each star is within the pattern. Table 6.1 shows the various patterning strategies and the frequency of participants who engaged in each strategy.

For five-year-olds in both conditions, a majority of children used pattern-relevant strategies in the pattern transfer task (10 out of 17 for the *exploration* condition and 9 out of 17 for the *demonstration* condition). A majority of six-year-old *demonstration* condition participants also used pattern-relevant strategies (11 out of 17). 7 of the 17 six-year-old *exploration* condition participants used the most adavanced exploit strategy, finding all three stars consecutively, compared to 4 of the 17 six-year-old *demonstration* condition participants. These preliminary results suggest that exploration may influence implicit pattern recognition and lead to better transfer and abstraction of patterns in new contexts.

6.1.3 Discussion

We provide preliminary evidence of the efficacy of utilizing exploration as a learning strategy in patterning for young children. This section discusses how exploration can further impact pedagogical methods and learning outcomes for children.

Exploration Encourages Active Engagement

Our results demonstrate that children can implicitly learn patterns even without direct instruction or feedback. As opposed to the *demonstration* condition, exploration enabled children to make sense of the pattern on their own. This active engagement may be a reason why we observed a trend of more *exploration* condition children engaging in an advanced, exploit strategy in the pattern transfer task. Because they were able to engage with the pattern previously, they may have been better able to connect the two structurally similar patterns.

Some children spontaneously explained their pattern reasoning during the pattern transfer task. For example, one five-year-old *exploration* participant said during the pattern transfer task, "*It looks like a pattern right? Listen, it's star, x, x, star, x, x, star, x, x.*" Similarly, a six-year-old in the *demonstration* condition referred to the pattern from right to left and explained, "*This is the pattern. It's no star, no star, star, no star, no star, star, no star, no star, star.*" Other children also pointed out the star and x pattern, but did so with respect to the shapes. For example, a five-year-old in the *exploration* condition stated, "*After two houses, one house has a star inside of it.*" A 6-year-old in the *exploration* condition similarly expressed, "*Wait, I realized something. There's a pattern: two houses that don't have a star and then one house that does have a star.*" Both children successfully found all the hidden stars using the advanced exploit strategy. Though not explicitly told the pattern to finding the hidden stars, both five- and six-year-old children demonstrated knowledge of the underlying pattern. More specifically, they showed understanding of the repeating pattern unit. Future interventions that utilize active engagement through exploration show promise for developing a general ability to identify repeating units

within a pattern even without direct instruction.

Our study examined the immediate learning effects of exploration. Future work could examine the longitudinal effects of using exploration as a learning strategy. For example, inventive activities that ask learners to produce original solutions prior to direct instruction prepare learners for future learning because they can connect their preconceptions to new knowledge gained [185]. Exploratory learning strategies may have a similar impact, priming children to pay attention to the high-level pattern structure rather than the specific components (*i.e.* shapes or stimuli) of the pattern.

Exploration Needs to Be Goal-Directed

Exploration as a problem-solving strategy can be inefficient, particularly when the goal itself may be ambiguous [62, 121, 205]. In our task, children had a clear goal to accomplish: find the hidden stars using a pattern. With this goal, children were directed in their exploratory efforts, which emerged through the strategies we observed. In addition, the task included immediate feedback. Children could see whether the house they selected contained a star underneath right after selection. This feedback may have helped children test and update their hypotheses of the pattern [197, 226]. While the pattern for finding the hidden stars was concrete, the differing strategies children used suggest that goal-directed exploration and immediate feedback may highlight alternative paths to solving the pattern. 28 of the 34 children in the *exploration* condition and 23 of the 34 children in the *demonstration* condition used some form of pattern-relevant strategy, demonstrating some understanding of the underlying pattern unit. In other forms of creative problem-solving, exploratory strategies should similarly enable people to quickly receive feedback and provide a clear goal to achieve through exploration.

Summary

We investigated whether exploration could help children learn and transfer patterns. In a between-subjects experiment, children either explored patterns themselves or watched a demonstration of the pattern. Our results suggest a trend that exploration may encourage more active engagement and highlight underlying structure. Future interventions that utilize exploration as a problem-solving strategy should provide mechanisms for immediate feedback and goal direction.

6.2   Problem-Framing Scaffolds: Hot & Cool Thinking in Creativity

Effective creative work requires both "hot" (exploratory) and "cool" (exploitative) thinking. Unfortunately, many people (especially novices) under-explore, jumping to the "cool" part too quickly, because they assume their current thinking "has to be" the path. This paper presents empirical results of how metaphorical problem framing scaffolds can influence creative performance. The task used De Bono's "Thinking Hats." In a between-subjects experiment comparing exploratory to exploitative problem frames, the exploratory problem frame led to more original designs and more diverse ideas during brainstorming. This work provides an empirical baseline

of how – even for short tasks – assigning people responsibility for broad thinking leads to better creative work.

6.2.1   Why are people so often averse to exploration?

Problem-solving engages a cycle of "hot" (exploratory) and "cool" (exploitative) thinking, searching broadly before narrowing onto any single idea or solution [120, 141](Figure 1.1). People often *satisfice*, stopping at the first solution they consider to be good enough [190]. Often, this is a savvy way to save time. However, the dark side of satisficing is when people underestimate the space of possibility, as often happens with complex creative problems, quickly settling on a path means forgoing (unconsidered) options that could be much better.

Strategies for Increasing Early Exploration Help to an Extent

To encourage broader exploration early in problem-solving, brainstorming, process models like the Double Diamond, and design thinking approaches encourage wild ideas at the outset. Despite this explicit encouragement, people – especially novices – still under-explore. The tendency to cool too soon and fixate can be difficult to overcome. Iteration and examples alone do not necessarily mean greater divergence as people tend to explore within a narrow solution space [49, 101, 127]. Without domain knowledge or strategies for exploration, people systematically underestimate how many better ideas are out there.

How Do We Overcome Satisficing?

We hypothesize that more explicitly attending to exploration and exploitation as phases may improve creativity. Approaches like brainstorming offer psychological safety and improve group's collective knowledge by "encouraging wild ideas" and "deferring judgment." One potential advantage of design thinking and related methods is that seeing problems from a user's perspective and drawing inspiration from existing practices and challenges gives people a different

vantage point from which to see ideas [46, 49, 199]. To *harness* better ideas, one must first see them. Despite compelling advances from leading practitioners, the empirical basis for stretching people's horizons remains limited.

This section explores the benefits of asking people to simply "look further," and also provides a more cognitive account of process than is available in the practitioner literature. We present a between-subjects study comparing an explore problem-framing scaffold to an exploit problem-framing scaffold. We found that with an exploratory framing, people provide more original designs and show greater diversity of ideas during brainstorming. This section demonstrates how process scaffolds connect to design outcomes and can "set the temperature" in creative thinking.

6.2.2   Experiment 1:   Can Problem-Framing Scaffolds Influence Creative Thought?

Method

This between-subjects experiment establishes a baseline of whether it is possible to influence creative outcomes by scaffolding how to frame problems. We adapt our scaffolds from DeBono's [46] Thinking Hats. These "hats" provide a metaphor for particular roles and purviews. DeBono's Green Hat represents creativity and serves as our explore scaffold (Figure 6.5a); the Blue Hat represents action and goal-setting and serves as our exploit scaffold (Figure 6.5b). We hypothesized that the explore scaffold would lead to more original ideas and broader search during the brainstorming period while the exploit scaffold would lead to more practical, but less original ideas and narrower search during the brainstorming period.

34 participants (27 female) were recruited from the Psychology & Cognitive Science subject pool (SONA) at a California research university. This task asked participants to redesign an aspect of the student eating experience to be better or more enjoyable. All participants received

Figure 6.5: a) The green thinking hat represents exploration, and b) the blue thinking hat represents exploitation. Participants in Experiment 1 used these scaffolds during brainstorming.

this scenario; half were randomly assigned to an *Explore* frame, half to an *Exploit* frame during the brainstorming period. The experiment gave five minutes for initial brainstorming and ten minutes to write a description of their preferred idea, why it is unique, and why it would improve the student eating experience. Lastly, a written survey asked participants questions about their experience with the Thinking Hat scaffolds.

Exploratory Framing Leads to More Original Designs & More Diverse Ideas

Two expert raters blind to condition with knowledge of campus issues and dining spaces rated each design on a 4-point Likert Scale for *originality, practicality*, and *justification* (how well motivated the design is). The inter-rater reliability between raters was substantial ($\kappa = 0.61$) Ratings were averaged between both raters for originality and practicality. *Explore* condition designs received significantly higher average ratings for originality ($m = 3.3, SD = 0.75$) than *Exploit* condition designs ($m = 2.4, SD = 0.88$)($F(1,48) = 15.2, R^2 = 0.22, p < .001$). Contrary to our hypothesis that the exploit scaffold would lead to higher practicality ratings, there were no significant differences between conditions for practicality ratings (*Explore m* $= 2.82, SD = 0.86$; *Exploit m* $= 3.02, SD = 0.93$)($F(1,48) = 0.62, R^2 = -0.01, p = .43$) between both conditions (Figure 6.6).

The expert raters also assessed the diversity of each participant's ideas on a 4-point Likert scale. Highly similar ideas (*i.e.* "make buffet-style payment system," "give students a
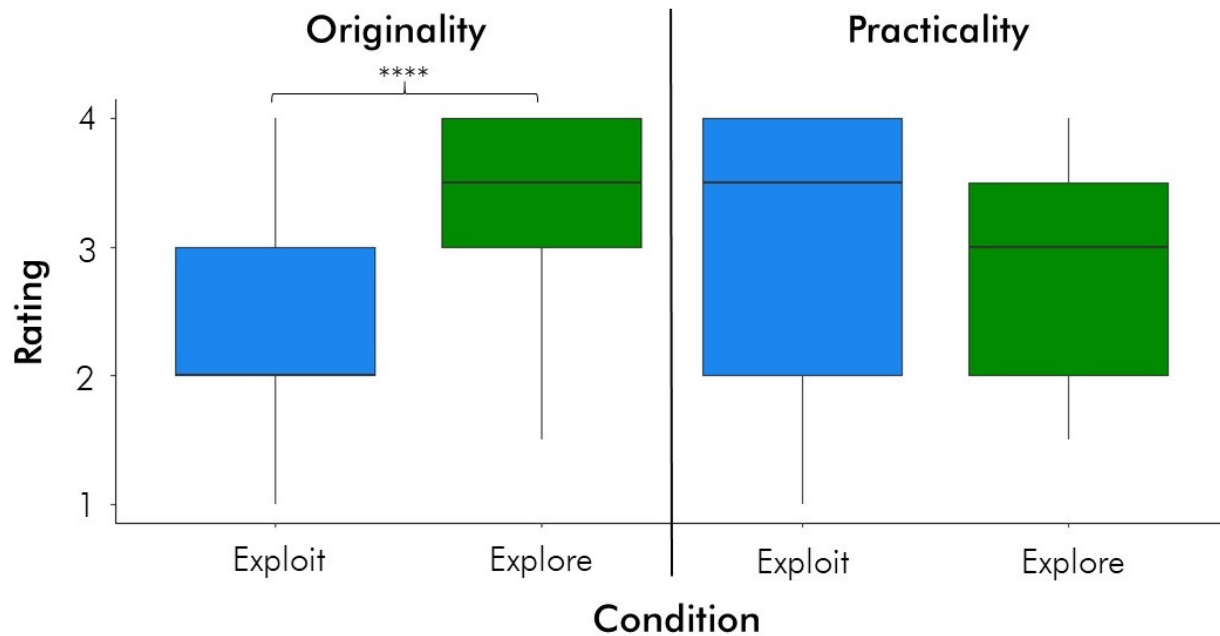
Figure 6.6: Participants with the *Explore* framing ($n = 17$) received significantly higher originality ratings for their designs than the *Exploit* framing ($n = 17$). There were no significant differences between practicality ratings. ****$F(1,48) = 15.2, R^2 = 0.22, p < .001$

certain number of swipes on payment cards") earned a 1 while highly diverse ideas (*i.e.* "chefs for students included in tuition," "grocery buddies/cooking mentors") earned a 4 on this scale. *Explore* participants ($m = 3.02, SD = 1.05$) brainstormed significantly more diverse ideas than *Exploit* participants ($m = 2.04, SD = 0.99$)($F(1,48) = 11.6, R^2 = 0.18, p < .005$) (Figure 6.7).

Exploratory Framing Enhances Perceptions of Creativity

The explore scaffold particularly helped participants push their boundaries of thought. One *Explore* participant stated that the without the Thinking Hat, *"I may have gotten to the same ideas..., but I probably would have refuted them on the spot with a reality check, or doubt."* Another stated that the scaffold *"made me question myself and think harder about the possibilities."* Two participants mentioned that the Thinking Hat as a metaphor was a particularly helpful metaphor. One participant said, *"The green hat had a placebo effect on me..., which influenced me to think outside the box and use my creativity regardless of feasibility."* Another

stated that, *"I could actually visualize wearing the hat [while] brainstorming new ideas."* One participant even said the thinking hat itself served as inspiration for his design.

Exploit Framing Encourages Real-World Action

Eight *Exploit* participants mentioned that the problem frame scaffold helped with organizing and structuring their thoughts towards goal-oriented solutions. One *Exploit* participant said the scaffold *"made me think about how [my ideas] could be practically accomplished, since it seemed to emphasize real-world action."* The exploit scaffold also seemed to be similar to participants' existing problem-solving strategies. Four participants believed their process of problem-solving without the thinking hat would have been the same. As one participant said, *"I think the blue thinking hat already describes my thinking process..., so I didn't have to constantly remind myself to think a certain way or use different methods to think."* None of the *Exploit* participants mentioned that the scaffold was challenging; they all felt that it added or supplemented
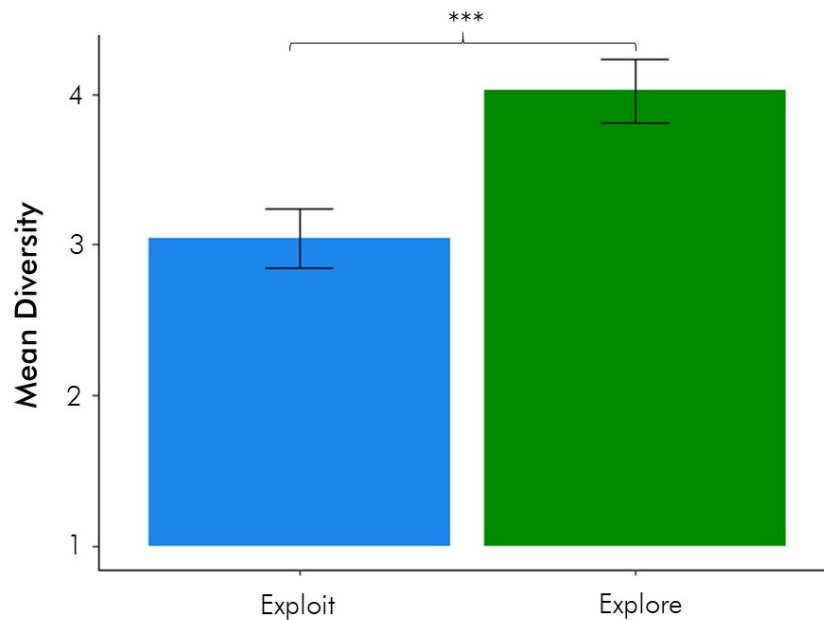
Figure 6.7: Participants with the *Explore* ($n = 17$) framing demonstrated greater diversity in their brainstormed ideas than those with the *Exploit* ($n = 17$) framing. ***$F(1,48) = 11.6, R^2 = 0.18, p < .005$

their ideation. However, one *Exploit* participant felt the framing limited his thinking, stating that it "*restricted him to analytical thinking versus being creative.*"

6.2.3 Experiment 2: Does the Order of Explore & Exploit Affect Creative Outcomes?

Method

A second experiment examined whether the order of the explore and exploit problem framing would impact creative outcomes. In this between-subjects study, 48 participants (30 female) from a California research university were recruited from an undergraduate design course. Similar to the first experiment, participants redesigned an aspect of the student eating experience. The task consisted of two initial brainstorming phases where participants used either explore or exploit scaffolds, the order of which was randomly counter-balanced between participants (23 participants in the *Explore-Exploit* condition, 25 participants in the *Exploit-Explore* condition). After each brainstorming phase, participants wrote a design idea, with the second design idea being the final design. Participants answered general questions about their experience in a post-survey after the experiment.

Explore Framing Improved Originality of First Design

Two experts in the food service industry blind to condition rated the final designs on two dimensions of *Originality* and *Practicality*, each on a 4-point Likert scale. Scores for each time point were averaged between both raters. The inter-rater reliability for raters was substantial ($\kappa = 0.69$). *Explore-Exploit* participants received higher average originality ratings for their first design than *Exploit-Explore* participants ($F(1,46) = 8.1, R^2 = 0.13, p < .01$). There were no significant differences in originality ratings between conditions for the second design ($F(1,46) = 1.08, R^2 = .002, p = .30$). There were also no significant differences in practicality
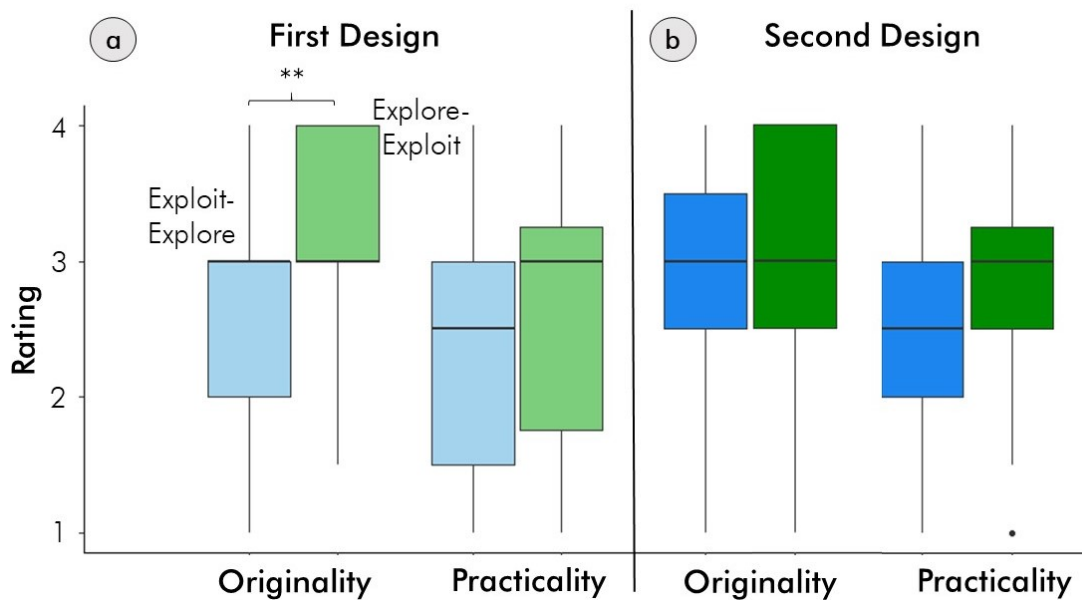
Figure 6.8: a) Participants with the *Explore-Exploit* order ($n = 23$) received higher originality ratings for their first designs than *Exploit-Explore* ($n = 25$) participants, but not differences in practicality ratings. b) We observed no significant differences in design outcomes for either measure in the second designs. A larger sample size or a change to the scaffolds themselves might result in more pronounced effects. **$F(1,46) = 8.1, R^2 = 0.13, p < .01$

ratings between conditions for both the first ($F(1,46) = 0.32, R^2 = -0.01, p = .58$) and second designs ($F(1,46) = 0.15, R^2 = 0.02, p = .70$)(Figure 6.8). As this work is preliminary, our results suggest a trend for further investigation.

Exploratory Scaffold Encouraged Wild Ideas

Despite our observed null quantitative results for most measures, the explore scaffold seemed successful in challenging participants to think more creatively based on qualitative reports. One participant in the *Explore-Exploit* condition noted, *"it pushed me to think of designs from different perspectives and points of view."* Similarly, one participant in the *Exploit-Explore* condition thought the explore scaffold *"opened the gates to better, more refined ideas."*

Some participants in both conditions thought the explore scaffold was challenging, but ultimately useful for creativity. For example, one *Explore-Exploit* participant said, *"it was*

shocking, but after the initial shock it helped spur a lot more creative ideas." A participant in the *Exploit-Explore* condition added, *"it forced me to step outside of my comfort zone and see things from a different perspective even if I didn't think that was possible."* This notion of feeling challenged may enhance creative thought regardless of order.

6.2.4   General Discussion & Next Steps

We provide preliminary empirical evidence of how problem framing can nudge people towards different cognitive processes and creative outcomes. This section addresses the implications and ongoing challenges of using this strategy.

Exploring Broadly is Difficult Without Prerequisite Domain Knowledge

The inclination to satisfice and find the "good enough" solution is particularly strong because of the time saved and effort required [190]. In both our experiments, the metaphor of an exploratory Thinking Hat seemed to enhance people's perceptions of their own creativity, which is a promising direction towards increasing productive exploration. Still, many final designs in Experiment 1 were similar to each other across conditions (for example, six designs mentioned sending out surveys for greater variety in dining halls on campus. Unfamiliarity with the domain space and the potential possibilities may have limited novices' ability to come up with fully original ideas.

Importantly, the explore scaffolds led to increased intra-exploration, where people's ideas during brainstorming were more diverse even if their final design outcome was similar to others. In Experiment 1, *Exploit* participants often came up with highly similar ideas or multiple examples of a single idea. In both experiments, participants frequently reported that the exploit scaffold was useful in structuring their thoughts, but it did not particularly challenge them to think differently. In contrast, many participants found the exploratory framing either challenging or effective in spurring new ideas. It may be that the perception or safety of feeling challenged

114

enhances originality. Additionally, wearing even a metaphorical hat may enable people to play out a "role" of exploration or exploitation, similar to prior work on human-centered mindsets [41, 199]. Our findings demonstrate that drawing attention to phase (explore or exploit) can influence creative mindset and potentially help in overcoming this tendency to satisfice and exploit early. An interesting area of future study might be examining how to further leverage this intra-exploration, perhaps by explicitly prompting for different perspectives or adding additional problem constraints.

Early Exploration Can Lead to Better Exploitation

Despite our hypothesis that a goal-oriented framing would lead to more practical ideas than the exploratory framing, we found no significant differences between the two conditions in both experiments. One potential concern of over-exploration is that wild ideas are produced without any being of practical use. Instead, broad exploration may help people realize the boundaries of a solution space without sacrificing practicality. We found this in Experiment 2 as well. Regardless of order, attending to exploration as a phase seemed fruitful in spurring more creative thought and ideas.

Exploration may also lead to more concrete and specific ideas. Four designs in the *Exploit* condition of Experiment 1 were overly broad. For example, one design suggested lowering the cost of food as a way to improve the student eating experience. While the design was well justified, it lacked specific details of how it could be put in place. An early exploration phase may indeed be necessary to refine concrete ideas later. This phase might attune people towards originality while maintaining their ability to later further develop and evaluate ideas. Future work should examine how to better scaffold the evaluation process to harness both unique and actionable ideas.

In Experiment 2, we observed that the *Explore-Exploit* condition produced more original first designs with *Exploit-Explore* participants improving slightly in Originality ratings for their

115

second design. This may suggest that prompting for exploration is beneficial regardless of where the person is in their creative process. Just as semantically diverse examples provide the most creative inspiration when a person reaches an impasse [30], prompts for exploration might be most useful when a person feels stuck or "out of ideas." Interestingly, a simple nudge, while not significant on its own, can at least catalyze getting unstuck.

What Scaffolds Improve Exploration?

Creative thinking combines domain knowledge with procedural strategies for greater exploration. Our experiments examine the latter, and our results suggest that how problems are framed affect people's thinking in different ways. Even simple metaphorical scaffolds like the Thinking Hats may be useful strategies in preparing novices to be creative by assigning a "role" to be creative. However, novices lack domain knowledge about a problem space and may need further help in understanding what creative means in a particular domain. To supplement this lack of knowledge, examples can provide anchors that usefully constrain a problem space [49, 127]. Pairing problem framing as an exploratory strategy with the use of examples may aid in eliciting more creative design outcomes. In addition, combining problem framing with more structured exploratory strategies, such as those provided at the Stanford dschool (www.dschool.stanford.edu), can be an effective pedagogical approach for creative learning.

Summary

We present empirical results of how the framing of problems can "set the temperature" of creative thinking. We found that an exploratory framing leads to more diverse ideas and more original design outcomes. The explore scaffold challenged participants to think more freely about a problem; the exploit scaffold helped participants structure and organize their thoughts. Future work should examine how such problem framing can be used with other strategies and scaffolds to catalyze creative learning.

Acknowledgements

This chapter, in part, is currently being prepared for submission for publication of the material by Tricia J. Ngoon, Vivian Leung, and Caren M. Walker. The dissertation author was the primary investigator and author of this material.

This chapter, in part, includes portions of material as it appears in *The Dark Side of Satisficing: Setting the Temperature of Creative Thinking* by Tricia J. Ngoon, Caren M. Walker, and Scott Klemmer in the Proceedings of the 2019 ACM Conference on Creativity and Cognition (C&C '19). The dissertation author was the primary investigator and author of this material.

Chapter 7

Conclusions & Future Work:

Opportunities for Fostering Creative

Learning

My hope is that this dissertation has demonstrated that attunement benefits creative exploration because knowing how to "see" is central to expertise. While simply telling people may not be sufficient (Section 6.2), I believe that richer attunement scaffolds can be highly effective. For example, this dissertation presented a simplification attunement scaffold to help people focus on high-level visual composition (Chapter 3), a scaffold for presenting conceptually-relevant examples (Chapter 4), and a scaffold for reusing high-quality examples in improving feedback (Chapter 5). This chapter describes opportunities for future work with these scaffolding strategies.

7.1   Embracing Ambiguity

I believe that a major roadblock to exploration and creative work is an emphasis on perfection or quantifiable metrics rather than iteration and progression. Creative work by nature is ambiguous, which makes fixating on the salient and the concrete all the more appealing, an "analysis paralysis" or choice paradox where so many options are possible that satisficing on one seems like the best solution to avoid being wrong [81, 181, 190]. Part of developing expertise

is learning what details to pay attention to and when and not getting lost among the 'trees.' For instance, focusing on the perfection of a character's outfit while sketching early on can distract (literally) from the bigger picture. Representations work best when the amount and type of expression they facilitate align well with the goal. One benefit of expertise is being able to break down the larger problems into simpler chunks where novices have greater difficulty in seeing how to do so. Abstraction blocks make concrete the intangible, abstract things that are easy to overlook. We found that groups using abstraction blocks in drawing focused more on composition concepts rather than details like what color a coat should be (Section 3.4). The key insight of abstraction blocks is that they make the ambiguous visible and tangible, emphasizing exploration rather than settling on the first known idea. Our study looked at abstraction blocks as an aid for chunking in drawing. Groups that used abstraction blocks formed composition plans with post-it notes as blocks (Figure 3.4). Abstraction blocks enabled top-down thinking, forming the plan before executing on the details. This draws similarities to backwards design in learning where the overall goal determines the plans and means to achieve that goal [224].

Setting high-level goals in creative work is difficult because the details distract us in two ways: 1) we do the small stuff to procrastinate from the tougher stuff, and 2) we get so focused on making the small stuff 'right' that we can't get past it. I believe the abstraction blocks approach can work domain-generally to help people set high-level goals by removing the details that distract and pigeonhole us. For example, in writing, focusing too heavily on grammar or wording early on can distract from the high-level narrative and flow of the writing [194]. Abstraction blocks for written work might first start out with blocks for general plot points (*i.e.* conflict and resolution or character development) before the writer begins composing sentences for each block. At an even higher level, what might abstraction blocks look like for making life decisions, like financial planning or figuring out career plans? Can these high-level goals be 'chunked' in a way that decomposes the details?

We can look to systems that already use a form of abstraction blocks. Scratch [169]
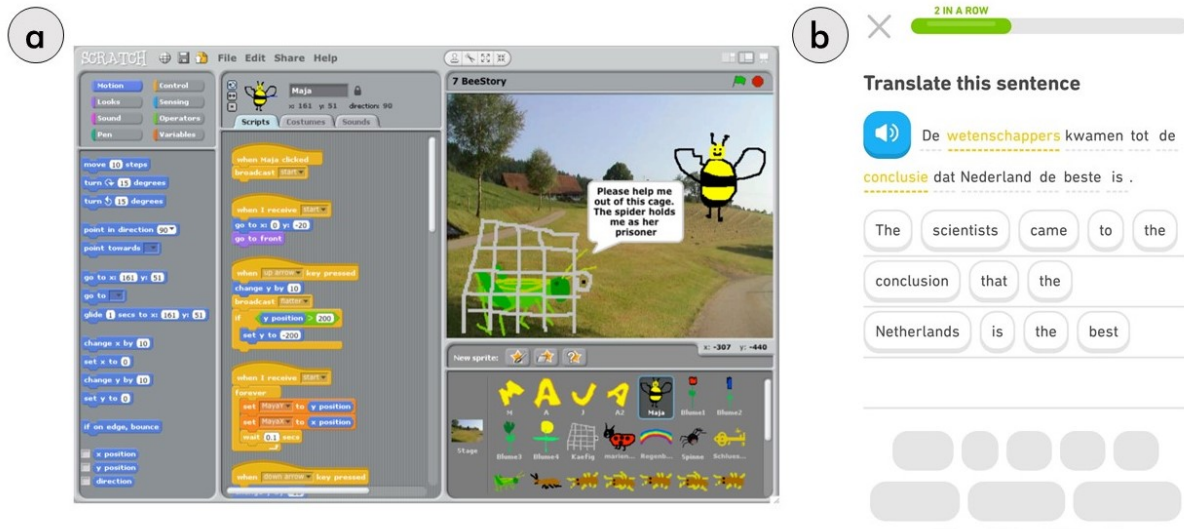
Figure 7.1: Similar to the concept of abstraction blocks, a) Scratch [169] uses block-based programming –short phrases visually rendered into discrete, parameterized blocks– to help novices easily create games and programs by understanding the higher-level concept without worrying about syntax or technical coding details. b) Duolingo [216](www.duolingo.com) uses blocks to help learners build and translate sentences without the need for minor spelling or grammar concerns.

and Duolingo [216] incorporate "blocks" as physical and moveable instantiations of coding or language learning to make learning more tangible (Figure 7.1). In our study, abstraction blocks made drawing almost like a puzzle, allowing groups to put the pieces together to create their picture. Finding ways to chunk problems into blocks and scaffolding the process of "assembling" them is a promising direction for creativity support tools. For example, tools could provide "levels," where users can toggle between more abstract, malleable chunks and more concrete detailed levels so they can easily shift between exploring the ambiguous and making concrete decisions on the details. Future work should examine how abstraction blocks might be applied as an oblique strategy for scaffolding attention to high-level plans and goals.

7.2 Meeting People Where They Are

The big vision in education more generally is taking less of a 'one-size-fits-all' approach and more of a personalized style to teaching and learning. Someone entering my Cognitive Science graduate program with a mathematics background is going to have a different knowledge base, skillset, and goal than someone with a psychology background, for instance, and the form of training and learning that works best for these two students will look different. The studio-apprentice model and one-on-one tutoring best exemplify this approach where help is adapted to individual learners [21, 34, 179]. In Schön's architecture scenario, the expert explains in detail the choices he makes to the student and engages in a Socratic dialogue [179]. Tailoring learning to one's individual knowledge and goals is tractable for the first time with computers. To emulate this dialogue, Shöwn's adaptive conceptual guidance presents general examples to users depending on their actions (Chapter 4). Shöwn used predetermined heuristics to decide what conceptual examples to show and when. What if tools could adapt the heuristics to users?

Every user has individual goals; some want to learn the concepts and gain deliberate practice in a skill while others might want to quickly execute a task without worrying about learning the concepts. Future creativity support tools should provide the most appropriate help for these goals. From our observations with Shöwn, we saw that proactive guidance gave novices an awareness of options to consider; user-led, reactive guidance was more useful for small, concrete tasks (Section 4.4). User-led searches are great when you know what to search for, but it's the unknown unknowns that make productive search both unlikely and extremely challenging if you don't know with what you need help. This is where a good tutor, whether a professional or peer, can proactively direct attention to issues you may not have considered and provide new insights. To close the loop, tutors can offer a record of the skills you have mastered and give targeted help for your specific areas of improvement. Similarly, a well-designed adaptive computational tool could proactively suggest help as you need it and create a record of skills you've mastered. Just

as a human tutor might give more or less help depending on a learner's progress, tools could also include different modes of help. For instance, a "maximum help" mode could provide the most guidance, both for conceptual understanding and technical skill help, while a "minimum help" or "silent" mode might provide only technical skill help or no help at all. These tools would give freedom to the user to decide which type of guidance they need depending on their own individual goals and tasks.

7.3   Creating Opportunities for Risk-Taking

Even today, many people hold a folk belief that mathematical abilities are innate rather than learned or grown [167]. Similarly, creativity seems to be another attribute that people think of as an innate talent or something accrued through deep experience in a given field rather than in a structured, deliberate manner. This dissertation argues that structured attunement can at the very least catalyze creative exploration and learning. Exploration inevitably includes giving up on some ideas. Most early ideas do not succeed, and the feeling of "wasting time" can limit further exploratory efforts. However, wild or even bad ideas can lead to serendipitous discoveries [159]. Creating low-stress and low-stakes environments can encourage exploration and risk-taking. Often, professional and educational settings engender an anxiety that makes people terrified to fail and therefore, terrified to take risks. Risk-taking requires that you can survive the failure.

Part of taking risks is understanding what aspects are worth the risk and what aspects need iteration. Feedback is a critical piece of this puzzle, learning what works and getting a direction for where to go [89, 174]. Our work with CritiqueKit showed that reusing exemplar feedback according to the structural critiera of being specific, actionable, and justified can improve feedback quality (Chapter 5). The next step is understanding how feedback is interpreted and applied. In human tutoring and studio models, experts use a variety of representations to explain concepts to a novice, such as sketches or deictic gestures like pointing [76, 179, 207]. To emulate how

an expert might provide feedback to a novice, particularly for visual or multimedia mediums, future feedback tools could allow reviewers to directly demonstrate their suggested improvements through multimodal and deictic interactions. Similar to how ReMap [61] uses deictic gestures and speech queries to find relevant moments in videos to present to users, feedback tools could also use these interactions as context to search for appropriate feedback suggestions from a corpus of critique examples. I hope the direction of critique tools help people both give feedback and better apply it to their work to give people the safety and freedom to take risks in their creative thinking.

7.4   A Future for Computational Coaching

A few years ago I began working with a boxing coach to improve my proficiency and technique with the goal of sparring more advanced students. My coach saw that despite my smaller size, I have quick footwork, sharp punches, and good handspeed. He tailored his coaching to improve my weaknesses in a way that capitalized on my strengths and showed me alternative ways to handle potential scenarios during a sparring session. Good coaches do exactly this: tailor their efforts to the individual learner and helping them hone their own style. I believe the future of creativity support tools is to emulate coaching, giving people the strategies and tools to accomplish their goals and improve. The future directions I've laid out in this chapter present opportunities to bring us closer to my vision. While this dissertation and other work within creativity support focus on digital learning and creation, an exciting area for computational tools and coaching is in physical domains like woodworking or dance. This dissertation presented three strategies that form the basis for what future computational coaching tools in any domain might become. These tools should provide attentional scaffolding to high-level concepts, adaptive support and relevant examples, and actionable feedback for improvement. We can learn skills and concepts from resources like video tutorials or books, but I believe that learning how to handle open-ended problems and ambiguous scenarios can come from adaptive and interactive guidance

Figure 7.2: Ohno! Has Talent is a choice-based assessment of critical thinking ability where students can explore different charts to decide which ones give the correct information for mixing the intended color (orange). If they create mix the colors correctly, Ohno! will begin singing. Figure adapted from [182].

from contextualized, personal coaching.

One promising avenue for adaptive coaching tools are those that assess motivation and choice. Choice-based assessments and activities examine what choices learners make such as whether they choose to explore resources or try different potential solutions [182]. Figure 7.2 shows an example of a 'choicelet' where students can choose to examine a catalog of color mixing charts to determine which charts are correct for their task as an assessment of critical thinking. Much like a human coach, computational coaches could help learners make and evaluate their choices could improve creative learning. In our study with Shöwn, we found that some participants in our study wanted to understand more about the reasoning behind suggestions and how the examples could apply to their own comics (Section 4.4). In addition to presenting suggested considerations, adaptive systems like Shöwn could also provide explanations for why those considerations might improve a user's work. Many creativity support systems have toolbars

that give users a variety of options of tools to user to execute their goals. What if coaching systems had "strategy bars" that allowed users to select different high-level strategies to use? This approach would give users a way to quickly explore different methods of achieving their goals and finding the strategy that works best for them.

Good coaches are few and far between, but the benefits of a coach are immeasurable. Issues of access and opportunity prevent some from even being able to afford to pursue their creative desires. Computational support tools could scale the talent of valuable coaches to nearly anyone who wants to learn. I believe that computational coaching tools should give opportunities to explore and connect exploration to their ultimate outcomes, meeting people where they are and helping them define and achieve their goals. I don't think computers should replace people's creativity, nor do I think they can. Rather, I see computational tools as a way of broadening people's horizons so they can accomplish more than they believe is possible.

7.5   Closing Remarks

Our information-driven society comprises an abundance of uncertainty, ambiguity, and information, in which we can often get lost in the woods. It is increasingly tempting to "cool" too soon, settling on what little certainty we can find. Without "seeing the forest for the trees," we may never come upon innovative solutions for our most complex problems. My dissertation introduces three strategies that help overcome this challenge. It contributes three interactive and contextual tools that embody these strategies to adaptively scaffold exploration at different stages of the creative process. It provides empirical evaluations and observations to demonstrate the efficacy of these strategies. Finally, the strategies introduced in this dissertation provide direction for open questions of fostering creative learning and scaling computational coaching to catalyze creativity for all. Ultimately, I envision a future in which our pedagogy and support tools empower people to practice everyday creativity to achieve their goals and expand human ambition.

Bibliography

[1] 2005. Swarm Sketch. (2005). http://swarmsketch.com/

[2] 2017. Gradescope. (2017). https://gradescope.com

[3] 2017. Turnitin Feedback Studio. (2017). http://turnitin.com/en

[4] Jessica Abel and Matt Madden. 2008. *Drawing Words and Writing Pictures: Making Comics: Manga, Graphic Novels, and Beyond*. Macmillan.

[5] Vincent Aleven. 2013. Help seeking and intelligent tutoring systems: Theoretical perspectives and a step towards theoretical integration. In *International handbook of metacognition and learning technologies*. Springer, 311–335.

[6] Vincent Aleven, Bruce McLaren, Ido Roll, and Kenneth Koedinger. 2004. Toward tutoring help seeking. In *International Conference on Intelligent Tutoring Systems*. Springer, 227–239.

[7] Vincent Aleven, Bruce M Mclaren, Jonathan Sewall, and Kenneth R Koedinger. 2009. A new paradigm for intelligent tutoring systems: Example-tracing tutors. *International Journal of Artificial Intelligence in Education* 19, 2 (2009), 105–154.

[8] Vincent Aleven, Ido Roll, Bruce M Mclaren, and Kenneth R Koedinger. 2010. Automated, Unobtrusive, Action-by-Action Assessment of Self-Regulation During Learning With an Intelligent Tutoring System. *Educational Psychologist* 45 (2010), 224–233. DOI: http://dx.doi.org/10.1080/00461520.2010.517740

[9] Vincent Aleven, Ido Roll, Bruce M McLaren, and Kenneth R Koedinger. 2016. Help helps, but only so much: Research on help seeking with intelligent tutoring systems. *International Journal of Artificial Intelligence in Education* 26, 1 (2016), 205–223.

[10] Vincent Aleven, Elmar Stahl, Silke Schworm, Frank Fischer, and Raven Wallace. 2003. Help seeking and help design in interactive learning environments. *Review of educational research* 73, 3 (2003), 277–320.

[11] Louis Alfieri, Timothy J Nokes-Malach, and Christian D Schunn. 2013. Learning through case comparisons: A meta-analytic review. *Educational Psychologist* 48, 2 (2013), 87–113.

[12] Andrew P Allen and Kevin E Thomas. 2011. A dual process account of creative thinking. *Creativity Research Journal* 23, 2 (2011), 109–118.

[13] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, and others. 2019. Guidelines for human-AI interaction. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–13.

[14] John R Anderson, Albert T Corbett, Kenneth R Koedinger, and Ray Pelletier. 1995. Cognitive Tutors: Lessons Learned. *Journal of the Learning Sciences* 4, 2 (1995), 167–207.

[15] Heidi Goodrich Andrade. 2005. Teaching with rubrics: The good, the bad, and the ugly. *College teaching* 53, 1 (2005), 27–31.

[16] Robert K Atkinson, Sharon J Derry, Alexander Renkl, and Donald Wortham. 2000. Learning from examples: Instructional principles from the worked examples research. *Review of educational research* 70, 2 (2000), 181–214.

[17] Yigal Attali. 2004. Exploring the Feedback and Revision Features of Criterion. In *The National Council on Measurement in Education (NCME)*.

[18] Ryan Shaun Baker, Albert T Corbett, and Kenneth R Koedinger. 2004. Detecting student misuse of intelligent tutoring systems. In *International conference on intelligent tutoring systems*. Springer, 531–540.

[19] Aditya Bharadwaj, Pao Siangliulue, Adam Marcus, and Kurt Luther. 2019. Critter: Augmenting Creative Work with Dynamic Checklists, Automated Quality Assurance, and Contextual Reviewer Feedback. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[20] Merim Bilalić, Peter McLeod, and Fernand Gobet. 2008. Why good thoughts block better ones: The mechanism of the pernicious Einstellung (set) effect. *Cognition* 108, 3 (2008), 652–661.

[21] Benjamin S Bloom. 1984. The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational researcher* 13, 6 (1984), 4–16.

[22] Sara A. Bly. 1988. A use of drawing surfaces in different collaborative settings. In *Proceedings of the 1988 ACM Conference on Computer-Supported Cooperative Work, CSCW 1988*. Association for Computing Machinery, Inc, 250–256. DOI: http://dx.doi.org/10.1145/62266.62286

[23] Elizabeth Bonawitz, Patrick Shafto, Hyowon Gweon, Noah D Goodman, Elizabeth Spelke, and Laura Schulz. 2011. The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition* 120, 3 (2011), 322–330.

[24] Michael Brooks, Sumit Basu, Charles Jacobs, and Lucy Vanderwende. 2014. Divide and Correct: Using Clusters to Grade Short Answers at Scale. In *Proceedings of the first ACM conference on Learning@ scale conference*. 89–98.

[25] Erin M Burgoon, Marlone D Henderson, and Arthur B Markman. 2013. There Are Many Ways to See the Forest for the Trees: A Tour Guide for Abstraction. *Perspectives on psychological science : a journal of the Association for Psychological Science* 8, 5 (sep 2013), 501–20. DOI:http://dx.doi.org/10.1177/1745691613497964

[26] William Buxton. 2007. *Sketching user experiences : getting the design right and the right design*. Morgan Kaufmann.

[27] Julia Cambre, Scott Klemmer, and Chinmay Kulkarni. 2018. Juxtapeer: Comparative peer review yields higher quality feedback and promotes deeper reflection. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.

[28] Joel Chan, Steven P Dow, and Christian D Schunn. 2018. Do the best design ideas (really) come from conceptually distant sources of inspiration? In *Engineering a Better Future*. Springer, Cham, 111–139.

[29] Joel Chan, Katherine Fu, Christian Schunn, Jonathan Cagan, Kristin Wood, and Kenneth Kotovsky. 2011. On the benefits and pitfalls of analogies for innovative design: Ideation performance based on analogical distance, commonness, and modality of examples. *Journal of mechanical design* 133, 8 (2011).

[30] Joel Chan, Pao Siangliulue, Denisa Qori McDonald, Ruixue Liu, Reza Moradinezhad, Safa Aman, Erin T Solovey, Krzysztof Z Gajos, and Steven P Dow. 2017. Semantically far inspirations considered harmful? accounting for cognitive states in collaborative ideation. In *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition*. 93–105.

[31] William G. Chase and Herbert A. Simon. 1973. Perception in chess. *Cognitive Psychology* 4, 1 (jan 1973), 55–81. DOI:http://dx.doi.org/10.1016/0010-0285(73)90004-2

[32] Min Chi, Ilsa Dohmen, Jonathan T Shemwell, Doris B Chin, Catherine C Chase, and Daniel L Schwartz. 2012. Seeing the forest from the trees: A comparison of two instructional models using contrasting cases. In *annual meeting of the American Educational Research Association, Vancouver, BC, Canada*.

[33] Michelene TH Chi, Paul J Feltovich, and Robert Glaser. 1981. Categorization and representation of physics problems by experts and novices. *Cognitive science* 5, 2 (1981), 121–152.

[34] Michelene TH Chi, Stephanie A Siler, Heisawn Jeong, Takashi Yamauchi, and Robert G Hausmann. 2001. Learning from human tutoring. *Cognitive science* 25, 4 (2001), 471–533.

[35] Michelene T H Chi. 1996. Constructing Self-Explanations and Scaffolded Explanations in Tutoring. *Applied Cognitive Psychology* 10 (1996), 33–49.

[36] Michelene T H Chi, Mirloi Bassok, Matthew W Lewis, Peter Reimann, and Robert Glaser. 1989. *Self-Explanations: How Students Study and Use Examples in Learning to Solve Problems*. Technical Report. 145–182 pages.

[37] Michelene T H Chi and Marguerite Roy. 2010. How Adaptive Is an Expert Human Tutor? *LNCS* 6094 (2010), 401–412.

[38] Pei-Yu Chi, Sally Ahn, Amanda Ren, Mira Dontcheva, Wilmot Li, and Björn Hartmann. 2012. MixT: automatic generation of step-by-step mixed media tutorials. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*. 93–102.

[39] Parmit K Chilana, Andrew J Ko, and Jacob O Wobbrock. 2012. LemonAid: selection-based crowdsourced contextual help for web applications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1549–1558.

[40] Kwangsu Cho, Christian D Schunn, and Davida Charney. 2006. Commenting on Writing: Typology and Perceived Helpfulness of Comments from Novice Peer Reviewers and Subject Matter Experts. *Written Communication* 23, 3 (2006), 260–294.

[41] Yung-Yi Juliet Chou and Barbara Tversky. 2017. Finding Creative New Ideas: Human-Centric Mindset Overshadows Mind-Wandering.. In *CogSci*.

[42] Yung-Yi Juliet Chou and Barbara Tversky. 2020. Changing perspective: Building creative mindsets. *Cognitive science* 44, 4 (2020).

[43] British Design Council. 2004. What is the framework for innovation? Design Council's Evolved Double Diamond. (2004). https://www.designcouncil.org.uk/news-opinion/what-framework-innovation-design-councils-evolved-double-diamond

[44] Nicholas Davis, Chih Pin Hsiao, Kunwar Yashraj Singh, Brenda Lin, and Brian Magerko. 2017. Creative sense-making: Quantifying interaction dynamics in co-creation. In *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition*. Association for Computing Machinery, Inc, 356–366. DOI:http://dx.doi.org/10.1145/3059454.3059478

[45] Nicholas Davis, Chih-Pin Hsiao, Kunwar Yashraj Singh, Lisa Li, and Brian Magerko. 2016. Empirically Studying Participatory Sense-Making in Abstract Drawing with a Co-Creative Cognitive Agent. In *Proceedings of the 21st ACM International Conference on Intelligent User Interfaces*. DOI:http://dx.doi.org/10.1145/2856767.2856795

[46] Edward De Bono. 1985. *Six thinking hats*. Little Brown and Company.

[47] Steven Dow, Julie Fortuna, Dan Schwartz, Beth Altringer, Daniel Schwartz, and Scott Klemmer. 2011. Prototyping dynamics: Sharing multiple designs improves exploration, group rapport, and results. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*. ACM Press.

[48] Steven P Dow, Alana Glassco, Jonathan Kass, Melissa Schwarz, Daniel L Schwartz, and Scott R Klemmer. 2010. Parallel prototyping leads to better design results, more divergence, and increased self-efficacy. *ACM Transactions on Computer-Human Interaction (TOCHI)* 17, 4 (2010), 1–24.

[49] Steven P Dow, Kate Heddleston, and Scott R Klemmer. 2009. The Efficacy of Prototyping Under Time Constraints. In *Proceedings of the seventh ACM conference on Creativity and cognition*. 165–174.

[50] Steven P Dow, Anand Kulkarni, Scott R Klemmer, and Björn Hartmann. 2012. Shepherding the Crowd Yields Better Work. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. 1013–1022.

[51] Karl Duncker and Lynne S Lees. 1945. On problem-solving. *Psychological monographs* 58, 5 (1945), i.

[52] Betty Edwards. 1979. *Drawing on the Right Side of the Brain* (4th ed.). Penguin Books Ltd.

[53] Will Eisner. 2008. *Comics and sequential art: Principles and practices from the legendary cartoonist*. WW Norton & Company.

[54] Douglas C Engelbart and William K English. 1968. A research center for augmenting human intellect. In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*. 395–410.

[55] K Anders Ericsson, Robert R Hoffman, and Aaron Kozbelt. 2018. *The Cambridge handbook of expertise and expert performance*. Cambridge University Press.

[56] K Anders Ericsson and Herbert A Simon. 1984. *Protocol analysis: Verbal reports as data*. the MIT Press.

[57] Ethan Fast, Colleen Lee, Alex Aiken, Michael S. Bernstein, Daphne Koller, and Eric Smith. 2013. Crowd-scale interactive formal reasoning and analytics. In *Proceedings of the 26th annual ACM symposium on User interface software and technology - UIST '13*. 363–372.

[58] John H. Flavell. 1979. Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist* 34, 10 (1979), 906–911. DOI: http://dx.doi.org/10.1037/0003-066X.34.10.906

[59] C. Ailie Fraser, Mira Dontcheva, Holger Winnemöller, Sheryl Ehrlich, and Scott Klemmer. 2016. DiscoverySpace: suggesting actions in complex software. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*. 1221–1232.

[60] C. Ailie Fraser, Joy O. Kim, Alison Thornsberry, Scott Klemmer, and Mira Dontcheva. 2019. Sharing the studio: How creative livestreaming can inspire, educate, and engage. In *Proceedings of the 2019 Creativity and Cognition*. Association for Computing Machinery, Inc, 144–155. DOI:http://dx.doi.org/10.1145/3325480.3325485

[61] C. Ailie Fraser, Julia M Markel, N James Basa, Mira Dontcheva, and Scott Klemmer. 2020. ReMap: Lowering the Barrier to Help-Seeking with Multimodal Search. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 979–986.

[62] C. Ailie Fraser, Tricia J Ngoon, Mira Dontcheva, and Scott Klemmer. 2019. RePlay: contextually presenting learning videos across software applications. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.

[63] Emily R Fyfe, Nicole M McNeil, and Bethany Rittle-Johnson. 2015. Easy as ABCABC: Abstract language facilitates performance on a concrete patterning task. *Child development* 86, 3 (2015), 927–935.

[64] Dedre Gentner. 1983. Structure-mapping: A theoretical framework for analogy. *Cognitive science* 7, 2 (1983), 155–170.

[65] Dedre Gentner. 2010. Bootstrapping the Mind: Analogical Processes and Symbol Systems. *Cognitive Science* 34, 5 (jul 2010), 752–775. DOI:http://dx.doi.org/10.1111/j.1551-6709.2010.01114.x

[66] Dedre Gentner, Sarah Brem, Ron Ferguson, and Philip Wolff. 1997. Analogy and creativity in the works of Johannes Kepler. (1997).

[67] Dedre Gentner, Jeffrey Loewenstein, and Leigh Thompson. 2003. Learning and transfer: A general role for analogical encoding. *Journal of educational psychology* 95, 2 (2003), 393.

[68] Graham Gibbs and Claire Simpson. 2004. Conditions Under Which Assessment Supports Students' Learning. *Learning and Teaching in Higher Education* 1, 1 (2004).

[69] Mary L Gick and Keith J Holyoak. 1980. Analogical problem solving. *Cognitive psychology* 12, 3 (1980), 306–355.

[70] Sarah Gielen, Elien Peeters, Filip Dochy, Patrick Onghena, and Katrien Struyven. 2010. Improving the effectiveness of peer feedback for learning. *Learning and Instruction* 20, 4 (aug 2010), 304–315. DOI:http://dx.doi.org/10.1016/J.LEARNINSTRUC.2009.08.007

[71] Elena L. Glassman, Aaron Lin, Carrie J. Cai, and Robert C. Miller. 2016. Learnersourcing Personalized Hints. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16*. ACM Press, 1624–1634.

[72] Elena L. Glassman, Jeremy Scott, Rishabh Singh, Philip J. Guo, and Robert C. Miller. 2015. OverCode: Visualizing Variation in Student Solutions to Programming Problems at Scale. *ACM Transactions on Computer-Human Interaction* 22, 2 (mar 2015), 1–35.

[73] Inga Glogger-Frey, Corinna Fleischer, Lisa Grüny, Julian Kappich, and Alexander Renkl. 2015. Inventing a solution and studying a worked solution prepare differently for learning from direct instruction. *Learning and Instruction* 39 (2015), 72–87. DOI: http://dx.doi.org/10.1016/j.learninstruc.2015.05.001

[74] Fernand Gobet, Peter C.R. Lane, Steve Croker, Peter C.H. Cheng, Gary Jones, Iain Oliver, and Julian M. Pine. 2001. Chunking mechanisms in human learning. (jun 2001). DOI: http://dx.doi.org/10.1016/S1364-6613(00)01662-4

[75] Fernand Gobet and Herbert A. Simon. 1998. Expert Chess Memory: Revisiting the Chunking Hypothesis. *Memory* 6, 3 (may 1998), 225–255. DOI:http://dx.doi.org/10.1080/741942359

[76] Susan Goldin-Meadow. 1999. The role of gesture in communication and thinking. *Trends in cognitive sciences* 3, 11 (1999), 419–429.

[77] Alison Gopnik, Shaun O'Grady, Christopher G Lucas, Thomas L Griffiths, Adrienne Wente, Sophie Bridgers, Rosie Aboody, Hoki Fung, and Ronald E Dahl. 2017. Changes in cognitive flexibility and hypothesis search across human life history from childhood to adolescence to adulthood. *Proceedings of the National Academy of Sciences* 114, 30 (2017), 7892–7899.

[78] Arthur C Graesser and Natalie K Person. 1994. Question Asking During Tutoring. *American Educational Research Journal Spring* 31, I (1994), 104–137.

[79] Arthur C. Graesser, Natalie K. Person, and Joseph P. Magliano. 1995. Collaborative dialogue patterns in naturalistic one-to-one tutoring. *Applied Cognitive Psychology* (1995). DOI: http://dx.doi.org/10.1002/acp.2350090604

[80] Arthur C Graesser, Kurt Vanlehn, Carolyn P Rosé, Pamela W Jordan, and Derek Harter. 2001. Intelligent Tutoring Systems with Conversational Dialogue. *AI Magazine* 22, 4 (dec 2001), 39–52. DOI:http://dx.doi.org/10.1609/AIMAG.V22I4.1591

[81] Adam Grant. 2021. *Think Again: The Power of Knowing What You Don't Know*. Penguin.

[82] Adam E. Green. 2016. Creativity, Within Reason. *Current Directions in Psychological Science* 25, 1 (feb 2016), 28–35. DOI:http://dx.doi.org/10.1177/0963721415618485

[83] Michael D. Greenberg, Matthew W. Easterday, and Elizabeth M. Gerber. 2015. Critiki: A Scaffolded Approach to Gathering Design Feedback from Paid Crowdworkers. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition - C&C '15*. 235–244.

[84] Tovi Grossman and George Fitzmaurice. 2010. ToolClips: an investigation of contextual video assistance for functionality understanding. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1515–1524.

[85] Nathan Hahn, Joseph Chang, Ji Eun Kim, and Aniket Kittur. 2016. The knowledge accelerator: Big picture thinking in small pieces. In *Conference on Human Factors in Computing Systems - Proceedings*. Association for Computing Machinery, 2258–2270. DOI:http://dx.doi.org/10.1145/2858036.2858364

[86] Björn Hartmann, Daniel MacDougall, Joel Brandt, and Scott R. Klemmer. 2010. What would other programmers do: suggesting solutions to error messages. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*.

[87] Giyoo Hatano and Kayoko Inagaki. 1986. *Two Courses of Expertise*. Technical Report.

[88] John Hattie. 2012. *Visible learning for teachers: Maximizing impact on learning*. Routledge.

[89] J. Hattie and H. Timperley. 2007. The Power of Feedback. *Review of Educational Research* 77, 1 (mar 2007), 81–112.

[90] Andrew Head, Elena Glassman, Gustavo Soares, Ryo Suzuki, Lucas Figueredo, and Loris D ' Antoni. 2017. Writing Reusable Code Feedback at Scale with Mixed-Initiative Program Synthesis. In *ACM Proceedings of Learning at Scale '17*. ACM.

[91] Marti Hearst. 2009. *Search user interfaces*. Cambridge university press.

[92] Jeffrey Heer. 2019. Agency plus automation: Designing artificial intelligence into interactive systems. *Proceedings of the National Academy of Sciences* 116, 6 (2019), 1844–1850.

[93] James W Hennessey, Han Liu, Holger Winnemöller, Mira Dontcheva, and Niloy J Mitra. *How2Sketch: Generating Easy-To-Follow Tutorials for Sketching 3D Objects*. Technical Report.

[94] Catherine M Hicks, Vineet Pandey, C. Ailie Fraser, and Scott Klemmer. 2016. Framing feedback: Choosing review environment features that support high quality peer assessment. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 458–469.

[95] Pamela J Hinds. 1999. The curse of expertise: The effects of expertise and debiasing methods on prediction of novice performance. *Journal of experimental psychology: applied* 5, 2 (1999), 205.

[96] Pamela J Hinds, Michael Patterson, and Jeffrey Pfeffer. 2001. Bothered by abstraction: The effect of expertise on knowledge transfer and subsequent novice performance. *Journal of applied psychology* 86, 6 (2001), 1232.

[97] Jim Hollan and Scott Stornetta. 1992. Beyond Being There. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, Vol. 92. 3–7.

[98] Forrest Huang and John F Canny. 2019. Sketchforme: Composing sketched scenes from text descriptions for interactive applications. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. 209–220.

[99] Hiroshi Ishii and Minoru Kobayashi. 1992. ClearBoard: A Seamless Medium for Shared Drawing and Conversation with Eye Contact. In *Proceedings of the SIGCHI conference on Human factors in computing systems*.

[100] E Jane L., Ohad Fried, Jingwan Lu, Jianming Zhang, Radomír Mech, Jose Echevarria, Pat Hanrahan, and James A Landay. 2020. Adaptive Photographic Composition Guidance.. In *CHI*. 1–13.

[101] David G Jansson and Steven M Smith. 1991. Design fixation. *Design studies* 12, 1 (1991), 3–11.

[102] Elahe Javadi. 2012. *The impact of user interface design on idea integration in electronic brainstorming: an attention-based view*. Ph.D. Dissertation. University of Illinois at Urbana-Champaign.

[103] Mads Møller Jensen, Roman Rädle, Clemens N Klokmose, and Susanne Bødker. 2018. Remediating a Design Tool: Implications of Digitizing Sticky Notes. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. DOI: http://dx.doi.org/10.1145/3173574.3173798

[104] Philip Nicholas Johnson-Laird. 1983. *Mental models: Towards a cognitive science of language, inference, and consciouness* (no. 6 ed.). Harvard University Press.

[105] Daniel Kahneman. 2011. *Thinking, fast and slow*. Macmillan.

[106] Shalini Kaleeswaran, Anirudh Santhiar, Aditya Kanade, and Sumit Gulwani. 2016. Semi-supervised verified feedback generation. In *Proceedings of the 2016 24th ACM SIGSOFT International Symposium on Foundations of Software Engineering - FSE 2016*. 739–750.

[107] Sean Kandel, Andreas Paepcke, Joseph Hellerstein, and Jeffrey Heer. 2011. Wrangler: Interactive visual specification of data transformation scripts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 3363–3372.

[108] Hyeonsu B Kang, Gabriel Amoako, Neil Sengupta, and Steven P Dow. 2018. Paragon: An online gallery for enhancing design feedback with visual examples. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.

[109] Manu Kapur. 2008. Productive Failure. *Cognition and Instruction* 26, 3 (jul 2008), 379–424. DOI:http://dx.doi.org/10.1080/07370000802212669

[110] Judy Kay, Peter Reimann, Elliot Diebold, and Bob Kummerfeld. 2013. MOOCs: So many learners, so much potential... *IEEE Intelligent systems* 28, 3 (2013), 70–77.

[111] David Kelley and Tom Kelley. 2013. *Creative confidence: Unleashing the creative potential within us all*. Crown Publishing Group, New York. 288 pages.

[112] Charles Kemp, Amy Perfors, and Joshua B Tenenbaum. 2007. Learning overhypotheses with hierarchical Bayesian models. *Developmental science* 10, 3 (2007), 307–321.

[113] Trina C Kershaw and Stellan Ohlsson. 2004. Multiple causes of difficulty in insight: the case of the nine-dot problem. *Journal of experimental psychology: learning, memory, and cognition* 30, 1 (2004), 3.

[114] Joy Kim, Justin Cheng, and Michael S Bernstein. 2014. Ensemble: exploring complementary strengths of leaders and crowds in creative collaboration. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. 745–755.

[115] Joy Kim, Mira Dontcheva, Wilmot Li, Michael S Bernstein, and Daniela Steinsapir. 2015. Motif: Supporting Novice Creativity through Expert Patterns. (2015). `DOI`: http://dx.doi.org/10.1145/2702123.2702507

[116] Juho Kim, Philip J Guo, Carrie J Cai, Shang-Wen Li, Krzysztof Z Gajos, and Robert C Miller. 2014a. Data-driven interaction techniques for improving navigation of educational videos. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*. 563–572.

[117] Joy Kim and Andŕes Monroy-Herńandez. 2016. Storia: Summarizing social media content based on narrative theory using crowdsourcing. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, Vol. 27. Association for Computing Machinery, 1018–1027. `DOI`:http://dx.doi.org/10.1145/2818048.2820072

[118] Juho Kim, Phu Tran Nguyen, Sarah Weir, Philip J Guo, Robert C Miller, and Krzysztof Z Gajos. 2014b. Crowdsourcing step-by-step information extraction to enhance existing how-to videos. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 4017–4026.

[119] Joy Kim, Sarah Sterman, Allegra Argent Beal Cohen, and Michael S. Bernstein. 2017. Mechanical novel: Crowdsourcing complex work through reflection and revision. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*. Association for Computing Machinery, 233–245. `DOI`:http://dx.doi.org/10.1145/2998181.2998196

[120] Scott Kirkpatrick, C Daniel Gelatt, and Mario P Vecchi. 1983. Optimization by simulated annealing. *science* 220, 4598 (1983), 671–680.

[121] Paul A. Kirschner, John Sweller, and Richard E. Clark. 2006. Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist* 41, 2 (jun 2006), 75–86. `DOI`:http://dx.doi.org/10.1207/s15326985ep4102$_1$

[122] David Kirsh. 2011. How marking in dance constitutes thinking with the body. *The External Mind. Perspectives on Semiosis, Distribution, and Situation in Cognition* (2011), 183–214.

[123] Scott R Klemmer, Björn Hartmann, and Leila Takayama. 2006. How bodies matter: five themes for interaction design. In *Proceedings of the 6th conference on Designing Interactive systems*. 140–149.

[124] Scott R Klemmer, Mark W Newman, Ryan Farrell, Mark Bilezikjian, and James A Landay. 2001. The designers' outpost: a tangible interface for collaborative web site. In *Proceedings of the 14th annual ACM symposium on User interface software and technology*. 1–10.

[125] Albert Koedinger, Kenneth R. & Corbett. 1997. Cognitive Tutors: Technology Bringing Learning Science to the Classroom. (1997).

[126] Markus Krause, Tom Garncarz, Jiaojiao Song, Elizabeth M. Gerber, Brian P. Bailey, and Steven P Dow. 2017. Critique Style Guide: Improving Crowdsourced Design Feedback with a Natural Language Model. In *Proceedings of CHI '17*.

[127] Chinmay Kulkarni, Steven Dow, and Scott Klemmer. 2012. Early and Repeated Exposure to Examples Improves Creative Work. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 34.

[128] Chinmay Kulkarni, Koh Pang Wei, Huy Le, Daniel Chia, Kathryn Papadopoulos, Justin Cheng, Daphne Koller, and Scott R Klemmer. 2013. Peer and self assessment in massive online classes. *ACM Transactions on Computer-Human Interaction (TOCHI)* 20, 6 (2013), 1–31.

[129] Chinmay E. Kulkarni, Michael S. Bernstein, and Scott R. Klemmer. 2015. PeerStudio: Rapid Peer Feedback Emphasizes Revision and Improves Performance. In *Proceedings of the Second (2015) ACM Conference on Learning @ Scale - L@S '15*. 75–84.

[130] Ranjitha Kumar, Jerry O Talton, Salman Ahmad, and Scott R Klemmer. 2011. Bricolage: example-based retargeting for web design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2197–2206.

[131] Stacey Kuznetsov and Eric Paulos. 2010. Rise of the expert amateur: DIY projects, communities, and cultures. In *Proceedings of the 6th Nordic conference on human-computer interaction: extending boundaries*. 295–304.

[132] James A Landay and Brad A Myers. 1996. Sketching storyboards to illustrate interface behaviors. In *Conference on Human Factors in Computing Systems: Conference companion on Human factors in computing systems: common ground*, Vol. 13. 193–194.

[133] Gierad P Laput, Mira Dontcheva, Gregg Wilensky, Walter Chang, Aseem Agarwala, Jason Linder, and Eytan Adar. 2013. Pixeltone: A multimodal interface for image editing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2185–2194.

[134] Brian Lee, Savil Srivastava, Ranjitha Kumar, Ronen Brafman, and Scott R Klemmer. 2010. Designing with interactive example galleries. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 2257–2266.

[135] Martin Lee and Daniel L Schwartz. 2009. Prospective Adaptation in the Use of External Representations. *Cognition and Instruction* 27, 4 (2009), 370–400. DOI: http://dx.doi.org/10.1080/07370000903221775

[136] Lan Li, Xiongyi Liu, and Allen L. Steckelberg. 2010. Assessor or assessee: How student learning improves by giving and receiving peer feedback. *British Journal of Educational Technology* 41, 3 (may 2010), 525–536. DOI:http://dx.doi.org/10.1111/j.1467-8535.2009.00968.x

[137] Wei Li, Justin Matejka, Tovi Grossman, Joseph A Konstan, and George Fitzmaurice. 2011. Design and evaluation of a command recommendation system for software applications. *ACM Transactions on Computer-Human Interaction (TOCHI)* 18, 2 (2011), 1–35.

[138] James Lin, Mark W Newman, Jason I Hong, and James A Landay. 2000. DENIM: Finding a tighter fit between tools and practice for web site design. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 510–517.

[139] Greg Little, Lydia B. Chilton, Max Goldman, and Robert C. Miller. 2010. Exploring iterative and parallel human computation processes. In *Workshop Proceedings - Human Computation Workshop 2010, HCOMP2010*. 68–76. DOI: http://dx.doi.org/10.1145/1837885.1837907

[140] Jeffrey Loewenstein, Leigh Thompson, and Dedre Gentner. 2003. Analogical learning in negotiation teams: Comparing cases promotes learning and transfer. *Academy of Management Learning & Education* 2, 2 (2003), 119–127.

[141] Christopher G Lucas, Sophie Bridgers, Thomas L Griffiths, and Alison Gopnik. 2014. When children are better (or at least more open-minded) learners than adults: Developmental differences in learning the forms of causal relationships. *Cognition* 131, 2 (2014), 284–299.

[142] Kurt Luther, Jari-Lee Tolentino, Wei Wu, Amy Pavel, Brian P. Bailey, Maneesh Agrawala, Björn Hartmann, and Steven P. Dow. 2015. Structuring, Aggregating, and Evaluating Crowdsourced Design Critique. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '15*. 473–485.

[143] Lena Mamykina, Linda Candy, and Ernest Edmonds. 2002. Collaborative creativity. *Commun. ACM* 45, 10 (2002), 96–99.

[144] Joe Marks, Brad Andalman, Paul A Beardsley, William Freeman, Sarah Gibson, Jessica Hodgins, Thomas Kang, Brian Mirtich, Hanspeter Pfister, Wheeler Ruml, and others. 1997. Design galleries: A general approach to setting parameters for computer graphics and animation. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*. 389–400.

[145] Richard L Marsh, Joshua D Landau, and Jason L Hicks. 1996. How examples may (and may not) constrain creativity. *Memory & cognition* 24, 5 (1996), 669–680.

[146] Justin Matejka, Tovi Grossman, and George Fitzmaurice. 2011. Ambient help. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2751–2760.

[147] Justin Matejka, Wei Li, Tovi Grossman, and George Fitzmaurice. 2009. CommunityCommands: command recommendations for software applications. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology*. 193–202.

[148] Scott McCloud. 2006. *Making comics: Storytelling Secrets of Comics, Manga, and Graphic Novels*. Harper Collins.

[149] Bruce M McLaren, Sung-Joo Lim, and Kenneth R Koedinger. 2008. When is assistance helpful to learning? Results in combining worked examples and intelligent tutoring. In *International conference on intelligent tutoring systems*. Springer, 677–680.

[150] Amir Shareghi Najar, Antonija Mitrovic, and Bruce M. McLaren. 2014. Adaptive Support versus Alternating Worked Examples and Tutored Problems: Which Leads to Better Learning? In *Lecture Notes in Computer Science 8538*. Springer, Cham, 171–182. DOI: http://dx.doi.org/10.1007/978-3-319-08786-3$_1$5

[151] Susanne Narciss and Katja Huth. 2006. Fostering achievement and motivation with bug-related tutoring feedback in a computer-based training for written subtraction. *Learning and Instruction* 16, 4 (aug 2006), 310–322. DOI: http://dx.doi.org/10.1016/J.LEARNINSTRUC.2006.07.003

[152] Allen Newell, J Clifford Shaw, and Herbert A Simon. 1962. The processes of creative thinking.. In *Contemporary Approaches to Creative Thinking, 1958, University of Colorado, CO, US; This paper was presented at the aforementioned symposium.* Atherton Press.

[153] Tricia J Ngoon, C. Ailie Fraser, Ariel S Weingarten, Mira Dontcheva, and Scott Klemmer. 2018. Interactive guidance techniques for improving creative feedback. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–11.

[154] Huy Nguyen, Wenting Xiong, and Diane Litman. 2016. Instant Feedback for Increasing the Presence of Solutions in Peer Reviews. *HLT-NAACL Demos* (2016), 6–10.

[155] David J Nicol and Debra Macfarlane-dick. 2006. Formative assessment and self-regulated learning : A model and seven principles of good feedback practice. *Studies in Higher Education (2006)* 31, 2 (2006), 199–218. DOI:http://dx.doi.org/10.1080/03075070600572090

[156] Changhoon Oh, Jungwoo Song, Jinhan Choi, Seonghyeon Kim, Sungwoo Lee, and Bongwon Suh. 2018. I Lead, You Help But Only with Enough Details: Understanding the User Experience of Co-Creation with Artificial Intelligence. (2018). DOI: http://dx.doi.org/10.1145/3173574.3174223

[157] Stellan Ohlsson. 1992. Information-processing explanations of insight and related phenomena. *Advances in the Psychology of Thinking* 1 (1992), 1–44.

[158] Gary M Olson and Judith S Olson. 2000. Distance matters. *Human–computer interaction* 15, 2-3 (2000), 139–178.

[159] Alex F Osborn. 1953. Applied imagination: Principles and Procedures of Creative Thinking. (1953).

[160] Marina Papic and Joanne Mulligan. 2007. The growth of early mathematical patterning: an intervention study. In *Mathematics: Essential Research Practice*, Vol. 2.

[161] Marina M Papic, Joanne T Mulligan, and Michael C Mitchelmore. 2011. Assessing the development of preschoolers' mathematical patterning. *Journal for Research in Mathematics Education* 42, 3 (2011), 237–268.

[162] Chris Piech, Jonathan Huang, Andy Nguyen, Mike Phulsuksombati, Mehran Sahami, and Leonidas Guibas. 2015. Learning Program Embeddings to Propagate Feedback on Student Code. In *International Conference on Machine Learning (ICML 2015)*.

[163] Henry Rankin Poore. 1967a. *Composition in Art*. Sterling Publishing Co.

[164] Henry Rankin Poore. 1967b. *Pictorial Compositon: An Introduction*. Sterling Publishing Co.

[165] Jill L. Quilici and Richard E. Mayer. 2002. Teaching students to recognize structural similarities between statistics word problems. *Applied Cognitive Psychology* 16, 3 (apr 2002), 325–342. DOI:http://dx.doi.org/10.1002/acp.796

[166] Arkalgud Ramaprasad. 1983. On the definition of feedback. *Behavioral Science* 28, 1 (jan 1983), 4–13. DOI:http://dx.doi.org/10.1002/bs.3830280103

[167] Aneeta Rattan, Catherine Good, and Carol S Dweck. 2012. "It's ok—Not everyone can be good at math": Instructors with an entity theory comfort (and demotivate) students. *Journal of Experimental Social Psychology* 48, 3 (2012), 731–737.

[168] Brian J Reiser. 2018. Scaffolding Complex Learning: The Mechanisms of Structuring and Problematizing Student Work. *THE JOURNAL OF THE LEARNING SCIENCES* 13, 3 (2018), 273–304.

[169] Mitchel Resnick, John Maloney, Andrés Monroy-Hernández, Natalie Rusk, Evelyn Eastmond, Karen Brennan, Eric Rosenbaum, Jay Silver, Brian Silverman, and Yasmin Kafai. 2009. *Scratch: Programming for All*. Technical Report 11. DOI: http://dx.doi.org/10.1145/1592761.1592779

[170] Daniela Retelny, Sébastien Robaszkiewicz, Alexandra To, Walter Lasecki, Jay Patel, Negar Rahmati, Tulsee Doshi, Melissa Valentine, and Michael S. Bernstein. 2014. Expert crowdsourcing with flash teams. In *UIST 2014 - Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*. Association for Computing Machinery, Inc, 75–86. DOI:http://dx.doi.org/10.1145/2642918.2647409

[171] Bethany Rittle-Johnson, Emily R Fyfe, Laura E McLean, and Katherine L McEldoon. 2013. Emerging understanding of patterning in 4-year-olds. *Journal of Cognition and Development* 14, 3 (2013), 376–396.

[172] Bethany Rittle-Johnson and Jon R Star. 2007. Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology* 99, 3 (2007), 561.

[173] Rod D. Roscoe, Laura K. Allen, Jennifer L. Weston, Scott A. Crossley, and Danielle S. McNamara. 2014. The Writing Pal Intelligent Tutoring System: Usability Testing and Development. *Computers and Composition* 34 (2014), 39–59.

[174] D Royce Sadler. 1989. Formative assessment and the design of instructional systems. *Instructional science* 18, 2 (1989), 119–144.

[175] Philip M Sadler and Eddie Good. 2006. The impact of self-and peer-grading on student learning. *Educational assessment* 11, 1 (2006), 1–31.

[176] Niloufar Salehi and Michael S. Bernstein. 2018. HIVe: Collective design through network rotation. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (nov 2018). DOI:http://dx.doi.org/10.1145/3274420

[177] Sarah Boxer. 2005. A United Nations of Weird-Looking Sketches - The New York Times. (2005).

[178] Arvind Satyanarayan and Jeffrey Heer. 2014. Authoring Narrative Visualizations with Ellipsis. *Computer Graphics Forum* 33, 3 (jun 2014), 361–370. DOI:http://dx.doi.org/10.1111/cgf.12392

[179] Donald A Schön. 1983. *The reflective practitioner: How professionals think in action*. Vol. 5126. Basic books.

[180] Christian D. Schunn, Mark U. McGregor, and Lelyn D. Saner. 2005. Expertise in ill-defined problem-solving domains as effective strategy use. *Memory and Cognition* 33, 8 (2005), 1377–1387. DOI:http://dx.doi.org/10.3758/BF03193370

[181] Barry Schwartz. 2004. *The paradox of choice: Why more is less*. Harper Collins.

[182] Daniel L Schwartz and Dylan Arena. 2013. *Measuring what matters most: Choice-based assessments for the digital age*. The MIT Press.

[183] Daniel L Schwartz, John D Bransford, David Sears, and others. 2005. Efficiency and innovation in transfer. *Transfer of learning from a modern multidisciplinary perspective* 3 (2005), 1–51.

[184] Daniel L. Schwartz, Catherine C. Chase, Marily A. Oppezzo, and Doris B. Chin. 2011. Practicing Versus Inventing With Contrasting Cases: The Effects of Telling First on Learning and Transfer. *Journal of Educational Psychology* 103, 4 (2011), 759–775. DOI:http://dx.doi.org/10.1037/a0025140

[185] Daniel L. Schwartz and Taylor Martin. 2004. Inventing to Prepare for Future Learning: The Hidden Efficiency of Encouraging Original Student Production in Statistics Instruction. *Cognition and Instruction* 22, 2 (jun 2004), 129–184.

[186] Burr Settles and Steven Dow. 2013. Let's get together: the formation and success of online creative collaborations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2009–2018.

[187] Oren Shapira, Nira Liberman, N Trope, and S Rim. 2012. Levels of mental construal. *Sage handbook of social cognition* (2012), 229–250.

[188] Ben Shneiderman. 2002. Promoting universal usability with multi-layer interface design. *ACM SIGCAPH Computers and the Physically Handicapped* 73-74 (2002), 1–8.

[189] Pao Siangliulue, Kenneth C Arnold, Krzysztof Z Gajos, and Steven P Dow. Toward Collaborative Ideation at Scale — Leveraging Ideas from Others to Generate More Creative and Diverse Ideas. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*. DOI:http://dx.doi.org/10.1145/2675133.2675239

[190] Herbert A Simon. 1972. Theories of bounded rationality. *Decision and organization* 1, 1 (1972), 161–176.

[191] Rishabh Singh, Sumit Gulwani, and Armando Solar-Lezama. 2013. Automated feedback generation for introductory programming assignments. *ACM SIGPLAN Notices* 48, 6 (jun 2013).

[192] Steven M. Smith, Thomas B. Ward, and Jay S. Schumacher. 1993. Constraining effects of examples in a creative generation task. *Memory & Cognition* 21, 6 (nov 1993), 837–845. DOI:http://dx.doi.org/10.3758/BF03202751

[193] Carolyn Snyder. 2003. *Paper Prototyping: The Fast and Easy Way to Design and Refine User Interfaces - Carolyn Snyder - Google Books*. Morgan Kaufmann.

[194] Nancy Sommers. 1980. Revision strategies of student writers and experienced adult writers. *College composition and communication* 31, 4 (1980), 378–388.

[195] Douglas Stone and Sheila Heen. 2015. *Thanks for the feedback: The science and art of receiving feedback well*. Penguin Books.

[196] Emily Sumner, Amy X Li, Amy Perfors, Brett Hayes, Danielle Navarro, and Barbara W Sarnecka. 2019a. The Exploration Advantage: Children's instinct to explore allows them to find information that adults miss. (2019).

[197] Emily Sumner, Mark Steyvers, and Barbara W Sarnecka. 2019b. It's not the treasure, it's the hunt: Children are more explorative on an explore/exploit task than adults.. In *CogSci*. 2891–2897.

[198] John C Tang. 1991. *Findings from observational studies of collaborative work*. Technical Report. 143–160 pages.

[199] Jaime Teevan and Lisa Yu. 2017. Bringing the Wisdom of the Crowd to an Individual by Having the Individual Assume Different Roles. In *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition*. ACM, 131–135.

[200] Michael Terry and Elizabeth D Mynatt. 2002. Side views: persistent, on-demand previews for open-ended tasks. In *Proceedings of the 15th annual ACM symposium on User interface software and technology*. 71–80.

[201] The Johnny Cash Project. 2012. the Johnny Cash Project. (2012). http://www.thejohnnycashproject.com/

[202] Maryam Tohidi, William Buxton, Ronald Baecker, and Abigail Sellen. 2006. Getting the right design and the design right. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. 1243–1252.

[203] K. Topping. 1998. Peer Assessment Between Students in Colleges and Universities. *Review of Educational Research* 68, 3 (jan 1998), 249–276.

[204] Sheng-Chau Tseng and Chin-Chung Tsai. 2007. On-line peer assessment and the role of the peer feedback: A study of high school computer course. *Computers & Education* 49, 4 (2007), 1161–1174.

[205] Juhani E Tuovinen and John Sweller. 1999. A comparison of cognitive load associated with discovery learning and worked examples. *Journal of educational psychology* 91, 2 (1999), 334.

[206] Amos Tversky and Daniel Kahneman. 1973. Availability: A heuristic for judging frequency and probability. *Cognitive psychology* 5, 2 (1973), 207–232.

[207] Barbara Tversky. 2011. Visualizing Thought. *Topics in Cognitive Science* 3 (2011), 499–535.

[208] Barbara Tversky and Paul U. Lee. 1999. Pictorial and Verbal Tools for Conveying Routes. Springer, Berlin, Heidelberg, 51–64. DOI:http://dx.doi.org/10.1007/3-540-48384-5$_4$

[209] Barbara Tversky and Masaki Suwa. 2009. Thinking with Sketches. In *Tools for Innovation*. Oxford University Press, 75–84. DOI: http://dx.doi.org/10.1093/acprof:oso/9780195381634.003.0004

[210] Melissa A. Valentine, Daniela Retelny, Alexandra To, Negar Rahmati, Tulsee Doshi, and Michael S. Bernstein. 2017. Flash organizations: Crowdsourcing complex work by structuring crowds as organizations. In *Conference on Human Factors in Computing Systems - Proceedings*, Vol. 2017-May. Association for Computing Machinery, 3523–3537. DOI: http://dx.doi.org/10.1145/3025453.3025811

[211] Hans van der Meij, Peter Blijleven, and Leanne Jansen. 2003. What makes up a procedure. *Content & Complexity. Information design in technical communication* (2003), 129–186.

[212] Hans van der Meij and Jan van der Meij. 2016. The effects of reviews in video tutorials. *Journal of computer assisted learning* 32, 4 (2016), 332–344.

[213] Kurt VanLehn. 1998. Analogy events: How examples are used during problem solving. *Cognitive Science* 22, 3 (1998), 347–388.

[214] Sara Värlander. 2008. The role of students' emotions in formal feedback situations. *Teaching in Higher Education* 13, 2 (apr 2008), 145–156. DOI: http://dx.doi.org/10.1080/13562510801923195

[215] Michael S Vendetti, Aaron Wu, and Keith J Holyoak. 2014. Far-out thinking: Generating solutions to distant analogies promotes relational thinking. *Psychological science* 25, 4 (2014), 928–933.

[216] Luis Von Ahn. 2013. Duolingo: learn a language for free while helping to translate the web. In *Proceedings of the 2013 international conference on Intelligent user interfaces*. 1–2.

[217] Edward Vul and Harold Pashler. 2008. Measuring the crowd within: Probabilistic representations within individuals. *Psychological Science* 19, 7 (2008), 645–647.

[218] Jagoda Walny, Bongshin Lee, Paul Johns, Nathalie Henry Riche, and Sheelagh Carpendale. 2012. Understanding pen and touch interaction for data exploration on interactive whiteboards. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2779–2788.

[219] Dakuo Wang, Judith S Olson, Jingwen Zhang, Trung Nguyen, and Gary M Olson. 2015. How students collaboratively write using google docs. *IConference 2015 Proceedings* (2015).

[220] Mark Ward and John Sweller. 1990. Structuring effective worked examples. *Cognition and instruction* 7, 1 (1990), 1–39.

[221] Molly McLure Wasko and Samer Faraj. 2005. Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *MIS quarterly* (2005), 35–57.

[222] Sarah Weir, Juho Kim, Krzysztof Z Gajos, and Robert C Miller. 2015. Learnersourcing subgoal labels for how-to videos. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 405–416.

[223] Malcolm Welch, David Barlex, and Hee Sook Lim. 2000. Sketching: Friend or foe to the novice designer? *International Journal of Technology and Design Education* 10, 2 (2000), 125–148.

[224] Grant P Wiggins, Grant Wiggins, and Jay McTighe. 2005. *Understanding by design*. Ascd.

[225] Jennifer Wiley. 1998. Expertise as mental set: The effects of domain knowledge in creative problem solving. *Memory & cognition* 26, 4 (1998), 716–730.

[226] Robert C Wilson, Elizabeth Bonawitz, Vincent D Costa, and R Becket Ebitz. 2021. Balancing exploration and exploitation with information and randomization. *Current Opinion in Behavioral Sciences* 38 (2021), 49–56.

[227] Beverly Park Woolf. 2010. *Building intelligent interactive tutors: Student-centered strategies for revolutionizing e-learning.* Morgan Kaufmann.

[228] W Xiong, D. Litman, and C. Schunn. 2012. Natural Language Processing techniques for researching and improving peer feedback. *Journal of Writing Research* 4, 2 (nov 2012), 155–176.

[229] Miao Yang, Richard Badger, and Zhen Yu. 2006. A comparative study of peer and teacher feedback in a Chinese EFL writing class. *Journal of Second Language Writing* 15, 3 (dec 2006), 179–200. `DOI`:http://dx.doi.org/10.1016/j.jslw.2006.09.004

[230] Dongwook Yoon, Nicholas Chen, Bernie Randles, Amy Cheatle, Corinna E Löckenhoff, Steven J Jackson, Abigail Sellen, and François Guimbretière. 2016. RichReview++ Deployment of a Collaborative Multi-modal Annotation System for Instructor Feedback and Peer Discussion. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing.* 195–205.

[231] Lixiu Yu, Aniket Kittur, and Robert E. Kraut. 2016. Encouraging "Outside-the-box" thinking in crowd innovation through identifying domains of expertise. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, Vol. 27. Association for Computing Machinery, 1214–1222. `DOI`:http://dx.doi.org/10.1145/2818048.2820025

[232] Alvin Yuan, Kurt Luther, Markus Krause, Sophie Isabel Vennix, Steven P Dow, and Björn Hartmann. 2016. Almost an expert: The effects of rubrics and expertise on perceived value of crowdsourced design critiques. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing.* 1005–1017.

[233] Li Yuan and SJ Powell. 2013. MOOCs and open education: Implications for higher education. (2013).