TOWARDS MIXED-INITIATIVE MACHINES FOR CREATIVE TASKS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Alexandra W. D. Bremers

May 2026

TOWARDS MIXED-INITIATIVE
MACHINES FOR CREATIVE TASKS

Alexandra W. D. Bremers, Ph.D.
Cornell University 2026

Fabrication machines are becoming more capable, yet most still operate as passive tools that wait for explicit commands. This dissertation investigates how fabrication machines might instead take initiative to support creative making, without automating away the enjoyment of the process. The central question is how to design machine initiative that is helpful and timely while preserving human agency, skill development, and the intrinsic satisfactions of craft.

The research proceeds in three parts. First, theoretical foundations are established, drawing on Sennett's analysis of craftsmanship, Schön's concept of reflective practice, and Horvitz's principles for mixed-initiative interaction. These foundations are then grounded in empirical understanding through interviews with maker entrepreneurs, observations of professional machine operators, and design explorations illustrating the design space for collaborative fabrication.

Second, methods are developed for observing and capturing the multimodal interactions that occur around fabrication machines. Through a study of 12 participants interacting with three tabletop machines (a clay extruder, a pen plotter, and a sewing machine), it is demonstrated that a minimal combination of egocentric and machine-mounted cameras captures the majority of relevant interaction information, and an instrumentation protocol is proposed. In the resulting FabriCam-5 dataset, implicit behavioral cues, such as hesitation, postural shifts, and gaze changes, are observed to precede explicit help-seeking,

suggesting that machines attending to these signals could intervene before makers reach the point of frustration.

Third, this premise is tested through a Wizard-of-Oz study comparing proactive and reactive machine assistance during a sewing task with 20 novice participants, resulting in the CoSew-4 dataset. Proactive assistance, in which the machine anticipated confusion based on behavioral cues and task stage, was perceived as significantly more timely than reactive assistance, with a large effect size, without increasing perceived annoyance. Proactive assistance also trended toward higher helpfulness ratings (see Section 5.5.1). These subjective differences appeared despite no measurable effect on task performance, suggesting that user experience measures may be more informative than efficiency metrics for evaluating co-creative systems. Study participants engaged socially with the machine and anticipated that their preferences for assistance would change as expertise developed.

This dissertation contributes: (1) resources for studying mixed-initiative interactions in creative tasks, including instrumentation methods, protocols, and two synchronized multi-view video datasets (FabriCam-5 and CoSew-4); (2) empirical evidence that proactive machine assistance improves novice makers' experience without undermining their engagement; and (3) design principles for mixed-initiative fabrication systems that balance technical support with the preservation of creative agency. Together, these contributions lay the groundwork that is required for the design and development of mixed-initiative machines for creative tasks.

BIOGRAPHICAL SKETCH

Alexandra Bremers obtained her Ph.D. in Information Science at Cornell Tech, with a minor in Computer Science. Her research addresses how interactive and intelligent machines can support human workflows during complex, physical, and creative tasks. Her work sits at the intersection of human-computer interaction, emerging technologies, and design, with a focus on building systems that empower people to create. She was a member of the founding task force that created the Cornell Department of Design Tech. Alongside research, Alexandra maintains a creative practice and has received recognition in design, writing, and art, with support from a Cornell PiTech AI in Arts & Culture Fellowship and from Cornell's Backslash art initiative.

Alexandra's academic background includes degrees in Artificial Intelligence (M.S., Utrecht University) and Industrial Design (B.S., Eindhoven University of Technology). During her doctoral studies, Alexandra completed research internships at Walt Disney Imagineering R&D (2025), Accenture Labs (2023), and Toyota Research Institute (2021). Prior to her Ph.D., she worked full-time as a Human-Machine Interface Researcher at Jaguar Land Rover in the United Kingdom, where she collaborated with the University of Cambridge (2017–2020). During this time, she was recognized through Oxford–Cambridge Rising Women in Science and Engineering, the JLR & Tata Innovista awards, and a Royal Commission for the Exhibition of 1851 Industrial Fellowship.

Voor Maarten

ACKNOWLEDGEMENTS

This dissertation would not exist without Wendy Ju, who has been incredibly supportive and inspiring through her never-ending infectious optimism, her hard work, and her sense of humor. She encouraged me to pursue my interests and was generous with her time and advice. It is an honor to study and learn from someone who has such a keen eye for interactions. Thank you, Wendy, for everything. Over the past 6 years, I have learned to see the world differently.

I am thankful to my committee members, François Guimbretière and Steve Marschner, for their time and thoughtful engagement with this work. François has been a true collaborator throughout; his insights around fabrication research sharpened my thinking at every stage, and many of the ideas in this dissertation are better because of his feedback. I also thank Malte Jung, who hosted me in his Robots in Groups lab during Fall 2021, giving me a home on the Cornell Ithaca campus. Thijs Roumen, Jorge Ortiz, and Nik Martelaro have been valued research collaborators, and I thank Antti Oulasvirta for hosting me in Helsinki.

Much of my research was shaped by collaborations with industry. I thank Celia Gomez and Michael Garber Barron for mentoring me during a fantastic internship experience at Walt Disney Imagineering and Disney Research in Glendale in the summer of 2025, with special thanks to Susan Harden, Hayley Darukhanavala, and Erick Moen for their warm welcome, support, and guidance. I was lucky to spend the summer of 2023 as an intern at Accenture Labs in San Francisco, where Alex Kass, Mike Kuniavsky, Mirjana Spasojevic, Manaswi Saha, Adolfo Ramirez-Aristizabal, and Alexandria Pabst made me feel like part of the team and pushed my work in new directions. Thanks go to Hiroshi Yasuda for a longstanding research collaboration which included a formative internship at the Toyota Research Institute in the summer of 2021. I am also grateful to Hirotsugu Kashimura at Amada and Stéphane Bouet at Nissan for industry collaborations which were great opportunities for learning, to Louisa Pancoast for sharing her expertise on dance, to the team at Evil Mad Scientist (now Bantam Tools) for their support around the AxiDraw, and to Vernier Metal for welcoming me for a site visit.

To my labmates and friends at Cornell Tech: Natalie Friedman, Frank Bu, David Goedicke, Ilan Mandel, Hauke Sandhaus, Teresa Parreira, Stacey Li, Matt Franchi, Albert Han, Mario Yang, Daniel Enriquez, Sean Hardesty Lewis, Julia Len, Saki Suzuki, Kyon Terao, Yuri Mukai, Dave Dey, Rei Lee, J.D. Zamfirescu-Pereira, JiHyun Jeong, Sharon Yavo-Ayalon, Sam Lee, Xy Fang, Ann-Kareen Gedeus, Sophie Bai, Tauhid Tanjim, Jonathan Segal, Promise Ekpo, Travers Rhodes, Chanwoo Chun, Niko Grupen, Nialah Wilson-Small, Yiran Zhao, and Amrit Kwatra. You made the long days shorter and the hard problems more fun. Thank you for the encouragement, the companionship, and the tea.

I was fortunate to advise several talented student interns whose work contributed to the explorations in this dissertation: Avital Dell'Ariccia, Tobias Weinberg, Itay Grinberg, Daniel Zhou, and Cooper Murr. Thank you for your energy, creativity, and hard work. I wish you all the best in your futures.

To my family: thank you for nourishing my curiosity from the start. I grew up enjoying an abundance of intellectual stimulation, knowing that no new hobby or interest was too crazy. Thank you for supporting my ambitions even if it meant that they took me to the other side of the world.

And to Maarten, who was there from the very beginning. I cannot believe that we actually pulled it off: two Ph.D.s on two continents, through a pandemic, at times driving through the desert or hiking among redwoods, at other times with an ocean between us. I know that together we can do anything.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Towards Mixed-Initiative Machines for Creative Tasks

> "There's always this joy about learning a new skill and picking up a new hobby. [...] A lot of people, when they feel fidgety, pick up knitting or crochet because it's something they can do with their hands and help them [calm down]." – Crocheter (male, 30s).

> "You know how in bowling there's bumpers? I wish there were that for throwing, to make it so I didn't mess up so much, because I felt like I spent so many hours trying to learn it. [But] I feel like there is definitely a pride in making something by myself." – Potter (female, 20s).

> "I truly hate not being able to intervene. That's not something I had with clay. I think that many people working in the world of clay now realize that there is value in being able to intervene mid-print. [...] [I see] something I could easily intervene in and save the whole print. But I can't do that because it's not in the code or embedded in the jet. It's not like I can take it out, play with it a little, put it back, and continue." – 3D printing expert (female, 20s).

These three makers[1] describe a tension that runs through creative work with machines. On one hand, they describe benefits of being able to use technologi-

[1] These quotes came from informal exploratory interviewing that took place prior to the work described in Chapter 3. They serve an illustrative purpose.

cal assistance while making. Yet, they also want to preserve the joy that comes from doing skilled work with their own hands, and retain an ability to intervene in the process. The question this dissertation addresses is how fabrication machines might provide that support, especially for novices, without automating away what makes making meaningful.

This question has become urgent as artificial intelligence transforms creative tools. Generative AI can now produce images, music, and text that once required years of human training. Similar capabilities are emerging for physical fabrication. The makers we interviewed distinguished between what a machine *could* do for them and what they *wanted* a machine to do. They valued the process of making: the embodied engagement, the skill development, and the satisfaction of solving problems all provided value beyond just the creation of a final product. Any machine designed to support creative work must carefully address this tension.

Mixed-initiative interaction offers a framework for thinking about this challenge. Rather than machines that passively execute commands or, at the other extreme, autonomously complete tasks, mixed-initiative systems allow both human and machine to take initiative depending on the situation [Horvitz, 1999]. One can imagine how a machine might notice confusion and offer guidance, even if the user might still decline that guidance and proceed independently. In mixed-initiative interaction, initiative flows to whoever is best positioned to act at each moment. For fabrication machines, this vision suggests systems that watch, wait, and intervene at appropriate moments, becoming more like attentive collaborators than automated tools.

Realizing this vision requires answering questions that cannot be resolved through speculation alone. What do makers want and need from collaborative machines? What would a machine need to sense to intervene appropriately? And does proactive machine assistance improve user experience, or does it feel intrusive? This dissertation provides empirical answers to these questions and lays the foundation for the development of future automated mixed-initiative machines for creative tasks.

1.2 Research Questions

This dissertation addresses one central research question:

> **How should fabrication machines take initiative to support creative making without undermining the intrinsic value of the making process?**

This question contains an inherent tension. Taking initiative means that the machine acts without being explicitly commanded, but acting at the wrong moment, or in the wrong way, could disrupt the flow and satisfaction that make creative work worthwhile. Resolving this tension requires understanding what makers value, what machines can sense, and how users respond to machines that take initiative. We address this central question through three sub-questions, each corresponding to a phase of the research:

> **Sub-question 1: What do makers need from collaborative machines, and what could mixed-initiative fabrication look like?**

Before building or testing concrete systems, we need to understand the design space. What are the theoretical foundations for thinking about machine initiative in creative contexts? What do makers themselves say about the balance between support and autonomy? What might collaborative fabrication machines look like? Chapter 2 addresses these questions through theoretical grounding, and Chapter 3 builds on this with interviews with makers, and iterative design explorations.

> **Sub-question 2: How can researchers observe and capture the interactions around fabrication machines that would inform the design of initiative-taking systems?**

A machine that takes initiative must eventually sense when intervention is appropriate. However, before we can design such sensing, we need methods for observing the rich, multi-modal interactions that occur around fabrication machines. What does instrumentation for this purpose look like? What can different camera perspectives reveal about human-machine fabrication interactions? And what can these observations tell us about opportunities for machine intervention? Chapter 4 addresses these questions through an investigation of multi-camera instrumentation across three fabrication machines.

> **Sub-question 3: Does proactive machine assistance improve user experience compared to reactive assistance, and what interaction patterns emerge?**

The preceding chapters establish what collaborative machines might look like and how to observe the interactions that would inform their design. Next, the question is whether users prefer machines that anticipate their needs. Chapter 5 addresses this question through a Wizard-of-Oz study with novices, comparing proactive and reactive assistance during a physical making task.

1.3 Thesis Statement

Creative making derives value from the process itself, from the embodied engagement, the skill development, and the satisfaction of solving problems. Fabrication machines are becoming increasingly capable of sensing and responding to material and process states. Yet these capabilities remain oriented toward the machine and the artifact, not the human maker. The ability to read a maker's hesitation, recognize confusion, or calibrate initiative to a maker's skill level and creative intent remains largely absent from fabrication machines available to makers. This dissertation investigates what it would take for fabrication machines to attend to the human side of collaboration. Through interviews with makers, systematic instrumentation of fabrication interactions, and a controlled comparison of proactive and reactive machine assistance during a sewing task, this work demonstrates that machines that anticipate a novice maker's needs are perceived more positively than those that wait to be asked, without increasing annoyance, and that makers value such support not only for its technical content, but also for the social and emotional presence it provides during the making process.

This argument rests on three lines of investigation. First, interviews with maker entrepreneurs and professional machine operators, together with iterative design explorations using pen plotters and Wizard-of-Oz systems, establish what makers value in the making process and illustrate the design space for collaborative fabrication (Chapters 2 and 3). Second, a multi-camera instrumentation study across three tabletop fabrication machines develops and validates a protocol for capturing the multi-modal interactions that would inform machine intervention design (Chapter 4). Third, a Wizard-of-Oz study comparing proactive and reactive machine assistance during a sewing task with 20 novice participants provides empirical evidence that anticipatory support enhances the subjective experience of human-machine collaboration without measurably affecting task performance (Chapter 5).

1.4 Approach and Overview

This dissertation tells one larger story across four chapters: understanding the theoretical underpinnings of collaborative machines, envisioning what collaborative fabrication machines could be, developing methods to observe the interactions that would inform their design, and testing whether proactive machine initiative improves user experience. The research was conducted inductively, beginning with concrete making practices, exploratory design work, and empirical observations, and moving toward broader questions and theoretical framing. The chapter structure reflects this trajectory: each chapter builds on what the previous one made visible, rather than testing propositions derived from theory in advance.

Chapter 2: Background establishes the theoretical foundations and design space for mixed-initiative fabrication. First, we draw on Sennett's analysis of craftsmanship to understand why the process of making matters, not just the product. Second, we use Schön's concept of reflective practice to articulate how skilled makers learn through a "conversation with materials" that automation might disrupt. Third, we build on Horvitz's principles for mixed-initiative interaction to frame how machines might take initiative while preserving human agency. These three theoretical schools of thought are then connected to recent literature in the space of fabrication research.

Chapter 3: Envisioning Collaborative Machines grounds theoretical foundations in empirical understanding of makers' experiences. Interviews with maker entrepreneurs reveal the tensions practitioners experience between wanting support and wanting to preserve the satisfactions of skilled making. Observations of professional machine operators illuminate how expertise and individual preference shape the desire for machine assistance. A series of design explorations with pen plotters and Wizard-of-Oz systems illustrates the design space for collaborative fabrication and identify relevant interaction dimensions.

Chapter 4: Capturing Interactions Around Machines develops the methodological foundation for studying human-machine fabrication. Through empirical investigation of 12 participants interacting with three tabletop fabrication machines (clay extruder, pen plotter, sewing machine), we demonstrate that a combination of egocentric and machine-mounted 360° cameras captures the majority of relevant interaction information. Applying a framework that attends to human, machine, artifact, and contextual state, we identify multiple breakdown types and find that behavioral cues indicating difficulty may often be visible before users explicitly seek help. The collected FabriCam-5 dataset will be made available via Harvard Dataverse.

Chapter 5: Prototyping Interactions Around Machines asks whether proactive machine assistance improves user experience compared to reactive assistance. Using Wizard-of-Oz methods, we compare two assistance conditions during a pillow plushie-sewing study with 20 novice participants. Proactive assistance was perceived as significantly more timely than reactive assistance, with a large effect size, without increasing perceived annoyance and despite no measurable differences in task performance. Qualitative analysis reveals that participants engaged with the machine socially and anticipated that their preferences would change with expertise. The chapter contributes the CoSew-4 dataset and articulates design implications for fabrication machines that balance initiative with human agency.

Chapter 6: Conclusion synthesizes findings across the empirical work to answer the central research question. We articulate design principles for mixed-initiative fabrication systems, including: timing matters as much as content; implicit signals may precede explicit requests; subjective experience diverges from objective efficiency; human-machine interaction is social and emotional; machine initiative should adapt to user expertise; and care must be taken to preserve human agency in creative decisions. We acknowledge limitations and propose directions for future research on co-creative fabrication machines.

1.5 Research Contributions

This dissertation makes three main contributions to the fields of Human-Computer Interaction, Human-Robot Interaction, and Digital Fabrication:

1. Methods for Studying and Prototyping Co-Creative Fabrication. This dissertation presents applied instrumentation approaches and two multi-modal datasets for investigating interactions around fabrication machines. The development and deployment of synchronized multi-camera recording systems, combined with the discussion of methods for analyzing implicit and explicit human-machine interactions, provides researchers with practical tools for studying fabrication workflows. The FabriCam-5 and CoSew-4 datasets provide synchronized multi-view video of fabrication interactions for further research. Together, these contributions lower barriers for researchers investigating co-creative machines.

2. Empirical Findings on Proactive Machine Assistance. Through controlled comparison of proactive and reactive assistance, this work demonstrates that anticipatory machine support is perceived as more timely and trends toward higher helpfulness ratings by novice users without increasing perceived annoyance (see Section 5.5.1). These findings extend mixed-initiative principles to physical fabrication contexts and show that the benefits of proactive assistance appear in user experience even when task performance shows no difference. The work also indicates that users engage socially with initiative-taking machines and that they expect their preferences to change as expertise develops.

3. Design Principles for Preserving Creative Agency in Automation. This dissertation advances a theoretical perspective on how automation should be designed for creative domains. The work articulates how machine initiative must be balanced against the intrinsic value humans derive from the creative process. These principles include: timing matters as much as content; implicit signals may precede explicit requests; subjective experience diverges from objective efficiency; human-machine interaction is social and emotional; machine initiative should adapt to user expertise; and care must be taken to preserve human agency. These principles provide guidance for designers of intelligent systems in creative contexts.

These contributions collectively establish a foundation for the next generation of fabrication machines as systems that can act as intelligent, responsive partners in creative work while preserving the agency, skill development, and joy that motivate people to create in the first place.

1.6 Prior Publications and Co-Authorship

Portions of this dissertation have appeared in prior publications with co-authors. The maker entrepreneur interview data discussed in Chapter 3 were collected as part of a larger collaborative project which led to the publication of Friedman et al. [2025][2]. The design explorations in Chapter 3 feature work and ideas that were published in Dell'Ariccia et al. [2022a,b], Grinberg et al. [2023], Bremers [2022a,b], Bremers and Ju [2024a,b][3][4]. Methodological descriptions of video interaction analysis in Chapter 4 were previously published in Bremers et al. [2024a][5]. The Wizard-of-Oz study in Chapter 5 builds on previous work on

[2] Data collection and thematic analysis were initially led by me, but the project was handed over to Natalie Friedman for finalizing analysis and leading the writing of the article. I remained actively involved in data analysis and writing.

[3] Avital Dell'Ariccia was a remote research intern who built the system for two Tic-Tac-Toe explorations and collected user study data. I supervised her, conceptualized the research project, decided the requirements of the system, designed the experiment, formed the argument of the articles, and edited the manuscripts.

[4] Itay Grinberg was a research intern who built an AxiDraw with extra degrees of movement, and a remote wizarding system. I supervised him, conceptualized the research project, decided the requirements of the system, conducted the co-creation of movement, formed the argument of the article, and edited the manuscript.

[5] See Bremers, A.W., Friedman, N., Lee, S., Wu, T., Laurier, E., Jung, M.F., Ortiz, J. and Ju, W., 2024, July. (Social) Trouble on the Road: Understanding and Addressing Social Discomfort in Shared Car Trips. In Proceedings of the 6th ACM Conference on Conversational User Interfaces (pp. 1-13).

task guidance systems, which was published in Bremers et al. [2024b][6]. My specific contributions to each project are noted in the relevant sections. While not discussed extensively in this dissertation, my concurrent work on affect-based error detection for human-robot interaction inspired the idea to bring implicit interactions to fabrication machines. This work has been published in Bremers et al. [2023a,b,c]. Where the works in Chapter 3 involved collaborators, this was noted. Chapters 4 and 5 consisted of work that was completed individually, but described in the first-person plural convention standard in the field of human-computer interaction.

[6] Shared first-authorship with Manaswi Saha, who advised my internship at Accenture Labs. I conducted the literature review, system development, study design, data collection, data analysis and initial writing. Manaswi edited the article after my internship had concluded.

CHAPTER 2

BACKGROUND

This chapter establishes the theoretical and scholarly foundations for investigating mixed-initiative fabrication machines. Before building or testing systems, we seek to understand the theoretical frameworks for thinking about machine initiative in creative contexts, and position this dissertation within the broader landscape of existing research. The chapter thus proceeds in two parts: we first establish theoretical foundations, and then we review related work on interactive fabrication and co-creative systems to position this dissertation within the field and identify the research gap it addresses (Section 2.2).

Chapter 3 then builds on these foundations with empirical work: interviews and observations that reveal what makers need from collaborative machines, and design explorations that illustrate the space of possibilities.

2.1 Theoretical Foundations

Three theoretical traditions inform this dissertation's approach to co-creative fabrication. First, Sennett's analysis of craftsmanship establishes why the process of making matters independently of its products. Second, Schön's concept of reflective practice describes how skilled practitioners learn through engagement with materials. Third, Horvitz's principles for mixed-initiative interaction, extended by Ju's framework for implicit interaction, provide foundations for designing systems where humans and machines share control. These traditions do not align perfectly; for example, Sennett is skeptical of automation while Horvitz is optimistic about machine assistance.

2.1.1 The Craftsman: Why Making Matters

Richard Sennett's seminal work *The Craftsman* [Sennett, 2008] provides a theoretical and philosophical foundation for understanding the nature of human processes of making. In this work, Sennett argues that craftsmanship represents a fundamental human impulse: the desire to do work well for its own sake. This impulse extends beyond traditional trades to any domain where practitioners develop skill through sustained engagement, including programming, medicine, and design. Crawford's *Shop Class as Soulcraft* [Crawford, 2009] makes a similar argument from the perspective of a motorcycle mechanic turned philosopher, stating that knowledge work has been overvalued relative to skilled trades. Csikszentmihalyi's work on flow [Csikszentmihalyi, 1990] documents the psychological rewards of skilled engagement as a state of absorbed concentration that emerges when challenge matches capability.

A central concept introduced by Sennett is that of *material consciousness*, which refers to intimate knowledge that is developed through sustained physical engagement with resistant materials. It is through years of practice that the craftsperson develops tacit knowledge that emerges from repetitive coordination between hand and eye, and this embodied understanding cannot be reduced to explicit rules or developed through verbal instruction alone. According to Sennett, this heightened sensitivity develops through an extended rhythm between perception and action. Research on skill acquisition supports this view. Studies of expert performance [Ericsson et al., 1993] emphasize the role of deliberate practice, i.e., sustained engagement with challenging tasks that provide immediate feedback. Embodied cognition research [Wilson, 2002] suggests that physical interaction with the world is not merely input to cognition but constitutive of it.

Specifically relevant for fabrication machines, Sennett warns against viewing automation as a substitute for skilled human labor, save for situations such as reducing dangerous tasks. In Sennett's view, skill development happens through physical presence and engagement with a material, guided by direct demonstration rather than instruction. He argues that the most valuable tools (i.e., *arousing tools*) are tools that stimulate imaginative engagement, rather than execute predetermined functions. For Sennett, augmentation that removes challenge also removes meaning.

However, one might say that Sennett's analysis could also be seen to romanticize manual work, underplaying its physical toll and economic precarity [Banks, 2014]. Not all repetitive manual labor produces the satisfactions Sennett describes; much of it is simply tedious. The maker movement that emerged in the 2000s [Dougherty, 2012] embraced technology more readily than Sennett, celebrating 3D printers and laser cutters as tools that democratize fabrication. Yet, even maker advocates acknowledge tension between digital fabrication and traditional craft values [Tanenbaum et al., 2013].

For this dissertation, Sennett's analysis raises a question: if material consciousness develops through physical engagement, what happens when machines mediate that engagement? A 3D printer that executes a digital file offers no resistance, no feedback, no opportunity for the rhythmic perception-action loop through which embodied knowledge develops[1]. This suggests that collaborative fabrication machines should preserve opportunities for material engage-

[1] Experienced 3D printer users might disagree and state that 3D printing is rarely a frictionless process, but we are here referring to the material engagement that is absent when a print successfully executes.

ment rather than eliminating them, but determining what "preserve" means in practice requires further investigation.

2.1.2 The Reflective Practitioner: Learning Through Making

Donald Schön's *The Reflective Practitioner* [Schön, 2017] critiques technical rationality (i.e., the notion that professional knowledge entails the straightforward application of scientific theory to practical problems). According to Schön, this model fails to acknowledge the uncertainty, uniqueness, and instability that practitioners encounter at work. Instead, he introduces the term *reflection-in-action*, which refers to what practitioners do when they are confronted with new situations. In these cases, the reflective practitioner performs experimental action in order to confront initial assumptions and construct new descriptions of the encountered phenomenon.

Schön describes design as a reflective conversation with the materials of a situation: the designer proposes moves, observes their consequences and adjusts subsequent actions based on how the situation at hand "talks back." The notion of reflective practice in design is fundamentally different from a conceptualization of design as a linear application of predetermined actions and solutions. As each design situation is unique, practitioners must discover the particular characteristics of their situation through engaged inquiry. Reflection-in-action differs from reflection-on-action (i.e., thinking about what one did after the fact), which cannot guide behavior in the moment. Reflection-in-action is thinking *while* doing, adjusting in real time.

Schön's framework has been influential in design theory. Cross [1982] builds on these concepts to articulate "designerly ways of knowing": modes of cognition specific to design practice. Lawson [2006] uses reflection-in-action to explain how designers navigate ill-defined problems. In HCI, Schön's ideas informed participatory design [Bødker and Grønbæk, 1996] and research-through-design methodologies [Zimmerman et al., 2007] that emphasize making as a form of inquiry.

Critics have questioned whether reflection-in-action is as spontaneous as Schön suggests. Eraut [1995] argues that much professional practice relies on routinized responses rather than in-the-moment reflection, and that Schön's examples selectively highlight reflective episodes. Dorst and Dijkhuis [1995] note that Schön's framework better describes experienced practitioners than novices, who may lack the repertoire of moves needed for fluid reflection-in-action.

For fabrication, Schön's analysis has direct implications. Conventional digital fabrication workflows separate design from making: the designer completes all decisions in CAD software and then sends a file to a machine that executes it. This separation prevents the reflective conversation. The material cannot "talk back" until fabrication completes. By then, it is too late to adjust. Goudswaard et al. [2024] document how this linear model conflicts with actual maker practice, which is iterative and exploratory (see Figure 2.3). Makers want to adjust mid-process, but conventional machines do not support this.

Collaborative fabrication machines might restore the reflective conversation by enabling real-time adjustment during fabrication. However, this requires the machine to participate in the conversation by sensing the situation, communicating its state, and responding to human input mid-process. The machine becomes not just a tool and executor, but a partner in reflection.

2.1.3 Mixed-Initiative Interaction: Sharing Control

Eric Horvitz's work "Principles of Mixed-Initiative User Interfaces" [Horvitz, 1999] establishes foundational concepts for designing systems that integrate automated services with direct manipulation. Horvitz addresses a debate in interface research: whether to focus on enhancing users' abilities to directly manipulate objects, or to develop automated agents that provide assistance. His response advocates for neither approach exclusively, but rather an elegant coupling of both.

Horvitz identifies challenges with purely automated approaches: poor inference of user goals, inadequate consideration of action costs and benefits, inappropriate timing, and insufficient mechanisms for users to guide or override automation. His twelve principles for mixed-initiative design address these challenges, emphasizing uncertainty management ("consider uncertainty about user goals," "scope precision to match uncertainty"), timing ("use timing sensitive to attention"), and user control ("enable efficient direct invocation and termination," "support collaborative refinement").

Subsequent work extended mixed-initiative interaction beyond desktop interfaces. Ju and Leifer's framework for implicit interaction [Ju and Leifer, 2008] extends mixed-initiative thinking to physical, situated systems, characterizing interactions along two dimensions: initiative (who drives the interaction) and attention (how much the interaction demands) (see Figure 2.1). Systems can be *proactive* (taking initiative) or *reactive* (waiting for commands), and they can be *foreground* (demanding attention) or *background* (minimizing attentional load); each of these axes represents a continuum. For skilled physical activity, back-

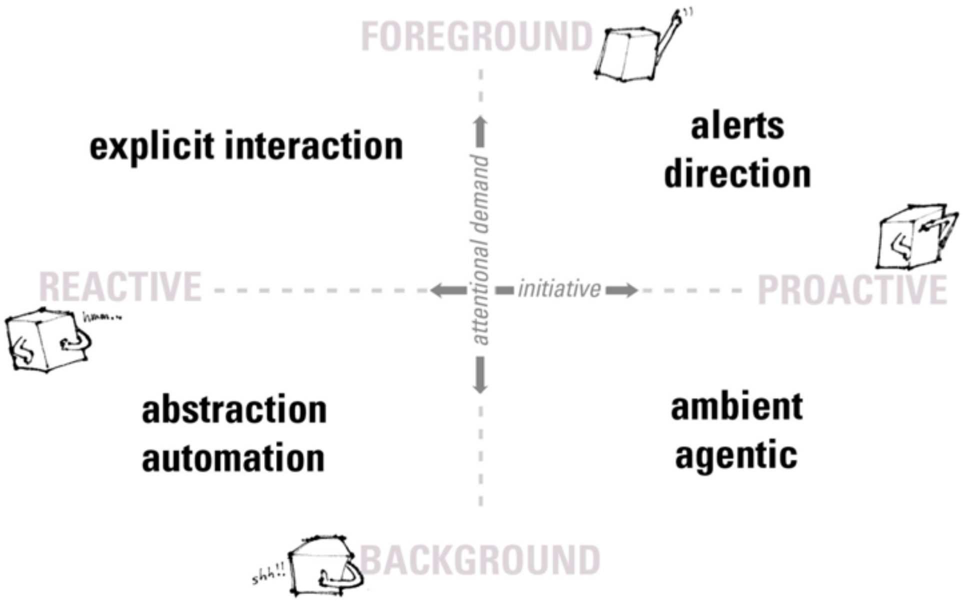


Figure 2.1: The Implicit Interaction Framework. (Image reprinted from Ju and Leifer [2008] with permission from the authors.)

ground interaction is particularly valuable as it could be assistive when needed, without interrupting the flow of making.

This framework suggests four quadrants of interaction style. Foreground-reactive systems (like traditional command-line interfaces) wait for explicit commands that require user attention. Foreground-proactive systems (like intrusive notifications) take initiative but demand attention. Background-reactive systems (like well-designed physical affordances) respond to implicit input without demanding focus. Background-proactive systems (like anticipatory climate control) take initiative while minimizing attentional demand. For fabrication, background-proactive may be ideal: machines that anticipate needs and provide assistance without disrupting makers' engagement with their work.

Empirical studies of initiative in physical collaboration reveal complexities. Mok et al. [2015] found that while proactive robots seemed more engaged, proactivity could negatively affect users' perception of their own status relative to the robot. Proactive assistance is not straightforwardly better; it carries social meaning. Users may interpret machine initiative as the machine claiming expertise or authority.

2.1.4 Synthesis: Navigating Theoretical Tensions

These three traditions do not align neatly. Sennett is skeptical of automation; Horvitz is optimistic about machine assistance. Schön emphasizes the practitioner's autonomy in the reflective conversation; mixed-initiative systems involve machines contributing to that conversation. Reconciling these perspectives requires distinguishing different forms of machine involvement.

Sennett's concern is automation that replaces human engagement: machines that execute tasks humans would otherwise perform, eliminate opportunities for skill development. Horvitz's vision is automation that augments human capability: machines contribute where they have comparative advantage while preserving human agency over goals and creative decisions. These are different phenomena: a machine that offers guidance to a confused user *augments*, while a machine that completes the task autonomously *replaces*. Schön's reflective conversation requires that actions have consequences the practitioner can perceive and respond to. Machines that mediate this conversation need not disrupt it and might even enrich it, providing feedback the material alone cannot give.

The synthesis that emerges is that collaborative fabrication machines should first of all augment rather than replace human engagement, preserving opportunities for skill development. They should support the reflective conversation by providing information and responding to human input, not by taking over decision-making, and take initiative judiciously, attending to timing and attentional demand. Furthermore, machines should respect uncertainty in user goals, particularly in creative contexts where goals emerge through making.

Whether these principles can be realized in practice is an open question; the design explorations in Chapter 3 and the studies in Chapters 4 and 5 contribute towards a foundation of this vision.

2.2 Related Work

This section reviews research on interactive fabrication and co-creative systems, positioning this dissertation within the field and identifying the research gap.

2.2.1 Interactive Fabrication

In "Interactive Fabrication: New Interfaces for Digital Fabrication", Willis et al. [2010] established a research agenda for real-time control of fabrication devices. They observe that while conventional CAD-coupled fabrication enables precise control, it removes makers from direct engagement with materials. *Interactive fabrication* explores when alternative interfaces for fabrication are useful, taking real-time input to design physical form not just before the fabrication process, but also while it is happening.

To this end, Willis et al. [2010] developed systems demonstrating various approaches to real-time fabrication control: Shaper, a device for interactive dispersing of foam material using a translucent touch interface directly above the fabrication workspace; Speaker, a wire sculpting device that uses sound input to visualize the corresponding sound wave in the form of the shaped wire; and Cutter, an interface that allows users to generate digital 3D models by physically manipulating a custom hotwire cutter. These systems aim to recover aspects of direct material engagement while retaining computational precision.

Subsequent research has extended interactive fabrication in multiple directions. For example, Mueller et al. [2013] developed Constructable, an interactive drafting table producing precise physical output through direct interaction. Users draw on workpieces using handheld laser pointers; the system tracks gestures, beautifies paths, and implements cuts using a high-powered laser. Mueller et al. [2014] introduced WirePrint to enable rapid iteration in 3D printing by fabricating wireframe structures. Peng et al. [2018] developed RoMA for interactive fabrication in augmented reality. These systems share a focus on enabling real-time human input during fabrication.

However, most present-day interactive fabrication systems position the human as the main or dominant source of initiative. The machine responds to human input but does not proactively offer assistance or anticipate needs. This dissertation explores the extension of interactive fabrication toward mixed-initiative systems where machines can take initiative alongside humans. Furthermore, it aims to investigate solutions that could be added to existing machines, rather than requiring the engineering of an entirely new machine.

2.2.2 Co-Creative Systems

Co-creative systems are designed to collaborate with humans on creative tasks, contributing ideas rather than just executing commands. Lin et al. [2020b] envisioned Cobbie, a mobile robot that iteratively ideates with designers through sketching. They found that physical robots were more satisfying than screen-based agents in motivating exploration. Jansen and Sklar [2021] identified ways collaborative drawing systems might assist creativity: corrective drawing, predictive drawing, scene completion, and artist-block mitigation. They also documented artists' skepticism toward automation of creative work.

Figure 2.2: CoFRIDA combines visual drawing input with text prompts to achieve human-robot collaborative drawing. (Image reprinted from Schaldenbrand et al. [2024], © 2024 IEEE.)

Recent physically co-creative systems have implemented mixed-initiative collaboration. Schaldenbrand et al. [2024] developed CoFRIDA, a framework for human-robot collaborative drawing where both parties respond to each other in real-time (see Figure 2.2). The system won Best Paper at ICRA 2024, demonstrating growing interest in physical co-creativity. However, CoFRIDA focuses primarily on semantic content (what is being drawn) rather than the timing and appropriateness of machine initiative (when the machine should act).

For fabrication specifically, Kim et al. [2017] presented design fictions proposing that collaborative fabrication machines need accessibility, fluidity, and concurrency. Goudswaard et al. [2024] described nonlinear human-machine collaborative 3D printing, emphasizing that real fabrication workflows rarely follow linear design-then-fabricate models. These works envision collaborative fabrication but do not empirically test how users respond to machines that take initiative.

Figure 2.3: Methods for designers to manipulate 3D printing during the making process (Image reprinted from Goudswaard et al. [2024] with permission from the authors.)

2.2.3 Machine Intelligence in Current Fabrication

While interactive fabrication and co-creative systems research focuses on new forms of human-machine collaboration, a parallel trajectory has made fabrication machines increasingly capable of sensing and responding to process conditions autonomously. This intelligence, however, is oriented toward the machine, the material, and/or the artifact rather than toward the human operator.

In additive manufacturing, real-time process monitoring includes Brion and Pattinson [2022]'s CAXTON (Collaborative Autonomous eXTrusiOn Network). CAXTON is a method that uses inexpensive webcams and multi-head neural networks to detect and correct diverse 3D printing errors in real time, generalizing across different geometries, materials, and printing systems. The system monitors material deposition through nozzle-focused cameras and automatically adjusts printing parameters (temperature, flow rate, speed, Z offset) when deviations from optimal conditions are detected. Commercial systems exist too: consumer 3D printers from manufacturers such as Bambu Lab now include camera-based first-layer inspection and automated failure detection that can pause prints when anomalies are identified. At the industrial scale, in-situ monitoring of tool condition and part wear can enable unmanned metal additive processes [Hassan et al., 2018].

Similar capabilities exist in other fabrication domains. In CNC machining, adaptive control systems have been developed to monitor tool condition and adjust cutting parameters in real time [Srinivasa Prasad et al., 2013, Hassan et al., 2018]. In sewing, research has produced intelligent systems that use neural networks and fuzzy logic to estimate fabric properties and automatically regulate thread tension during stitching [Koustoumpardis and Aspragathos, 2014]. Commercial sewing machines increasingly feature automatic thread tension, adaptive presser foot pressure, and computerized stitch regulation.

What unites these developments is their focus: they sense machine state (tool wear, nozzle temperature, motor current), material state (fabric tension, material behavior, cutting resistance), and artifact state (dimensional accuracy, surface quality, stitch formation). They respond with parameter adjustments or process corrections aimed at maintaining fabrication quality. Yet, they do not sense the human maker's experiential state: whether the operator is confused, whether a novice needs encouragement, whether the user's creative intent has shifted, or whether intervention would be welcome at a given moment.

2.2.4 Research Gap

Existing research thus leaves a gap at the intersection of two trajectories. On one hand, interactive fabrication and co-creative systems demonstrate the value of real-time human-machine collaboration during making, but position the human as the sole source of initiative. On the other hand, intelligent fabrication systems demonstrate that machines can sense, reason about, and respond to process conditions autonomously, but orient this intelligence toward the machine, material, and artifact rather than toward the human maker.

The underexplored area is where a system combines these capabilities: a fabrication machine that attends to the human side of the interaction, sensing when a maker is struggling, recognizing opportunities to offer guidance, and calibrating its initiative to the user's skill level and creative intent. Realizing such a system requires answering questions that neither tradition has addressed. What do makers need from machines that take initiative? How can researchers observe the multi-modal interactions around fabrication machines to inform the design of such systems? And does proactive machine assistance improve user experience, or does it feel intrusive?

This dissertation addresses the gap by investigating mixed-initiative interaction in physical fabrication contexts. We ask not only what collaborative machines might look like (see Section 3.2 and Section 3.3) but how users respond when machines take initiative during making tasks. This requires understanding what makers need (Chapter 3), developing methods to observe fabrication interactions (Chapter 4), and empirically comparing different initiative levels (Chapter 5).

CHAPTER 3

PART I: ENVISIONING COLLABORATIVE MACHINES

Chapter 2 established theoretical foundations for thinking about machine initiative in creative contexts. This chapter addresses the first sub-question of this dissertation: What does mixed-initiative fabrication look like, and what do makers need from collaborative machines? Before building or testing systems, we sought to understand the design space for collaborative fabrication: what theoretical foundations apply, what makers themselves experience when working with machines, and what collaborative fabrication might look like in practice.

This chapter proceeds in three parts. We first present empirical work to understand makers' experiences through interviews with maker entrepreneurs and observations of professional machine operators (Section 3.1). The interviews revealed what makers value about the process of making and what challenges they face. Second, we discuss three interaction vignettes that envision different roles a collaborative machine might take. Third, we describe a series of design explorations that illustrate the design space for collaborative fabrication machines (Section 3.3). The work presented here was conducted iteratively over several semesters; interviews, site visits, and design explorations informed each other throughout. The overall approach was inductive: rather than beginning with theory and testing propositions, this research moved from concrete making practices and exploratory design work toward broader questions and theoretical framing. Chapter 2 presents the theoretical foundations that emerged from and were refined through this process; they are introduced first to orient the reader, though in practice theory and empirical work developed together. Content in this chapter is therefore presented thematically rather than chronologically, moving from empirical understanding of makers' needs to the design dimensions those needs suggest.

The design explorations in Section 3.3 are positioned within the tradition of Research through Design [Zimmerman et al., 2007, Gaver, 2012]. Zimmerman et al. [2007] propose Research through Design as a legitimate research method for HCI, in which designers engage wicked problems by focusing on making artifacts that envision and work toward a preferred state of the world. In this model, the designed artifacts are evaluated for their process, invention, relevance, and extensibility. Gaver [2012] argues that rather than striving for extensible and verifiable theory, Research through Design should focus on its strength in producing theories that are provisional, contingent, and aspirational. Gaver [2012] further states that the appropriate role of theory in this tradition is to *annotate* collections of design examples, resulting in an annotated portfolio that illustrates a design space. Our explorations are also grounded in what Cross [1982] calls designerly ways of knowing: a mode of inquiry distinct from scientific or humanistic reasoning, in which knowledge resides in both the processes and the products of designing. Following Buxton [2010]'s distinction between sketching and prototyping, the explorations here are design sketches rather than prototypes. Sketches are quick, tentative, and deliberately incomplete; their purpose is to explore the problem space and ask what to build, not to evaluate how well a finished system works. Taken together, the five explorations in this chapter constitute a Research through Design contribution in the sense Zimmerman and Gaver describe: a set of design moves, each of which transforms what the next move could be, and which collectively make visible the design space for collaborative fabrication.

As such, this chapter contributes: (1) empirical observations on what makers value about the process of making and what challenges they face when working with machines, drawn from interviews with maker entrepreneurs and a site visit with professional machine operators; (2) three interaction vignettes (guide, companion, adaptive collaborator) that frame different possible relationships between makers and collaborative machines; and (3) a series of design sketches using pen plotters, Wizard-of-Oz systems, and cardboard mock-ups that explore five dimensions of the design space for collaborative fabrication: turn-taking, shared authorship, real-time collaboration, expressive movement, and physical scale. Together, these establish the design vision and open questions that motivate the empirical studies in Chapters 4 and 5.

3.1 Understanding Makers

The theoretical foundations and related work establish frameworks for thinking about co-creative fabrication, but frameworks must be grounded in empirical understanding of what makers experience. This section draws on two sources: an interview study with maker entrepreneurs (see Figure 3.1), and an observational site visit with professional machine operators at a sheet metal fabrication workshop (see Figure 3.2).

3.1.1 Maker Entrepreneurs

To understand the experiences of people who make physical goods, we conducted an interview study with maker entrepreneurs, i.e., individuals who de-

Figure 3.1: Products made by interviewed maker entrepreneurs. Top left: P27's hair clip; Top middle: P18's 3D printed stickers; Top right: P26's cardboard cuffs; Bottom left: P25's cut gem; Bottom middle: P28's wire sculptures; Bottom right: P19's Featherwing; Rightmost: P29's guitar tennis rackets. (For more information, see Friedman et al. [2025].)

sign, produce, and sell handcrafted items such as pottery, jewelry, and textiles [Friedman et al., 2025][1]. The study involved 20 USA-based maker entrepreneurs, six creative service entrepreneurs, and seven support personnel. While the primary focus was understanding opportunities for technology to support business aspects of making, the data revealed insights about the relationship between making, technology, and wellbeing. Interviews were semi-structured and analyzed using thematic analysis [Friedman et al., 2025]. We draw here on the subset of findings concerning makers' relationships with their tools, materials, and the process of making.

Most of the makers were "accidental entrepreneurs" who did not set out to start businesses but found themselves selling work to sustain their practice. As

[1] The main publication of this study was first-authored by Natalie Friedman, but I was a collaborator focused on experience of making throughout the project from beginning to end. Quotes are excerpted from Friedman, N., Bremers, A., Nyanyo, A., Clark, I., Kotturi, Y., Dabbish, L., Ju, W. and Martelaro, N., 2025. Understanding the challenges of maker entrepreneurship. Proceedings of the ACM on Human-Computer Interaction, 9(2), pp.1-29. [Friedman et al., 2025]

one gem cutter explained, she began selling gems so she could continue cutting while maintaining her hobby in a financially sustainable way. This identity as maker-first, entrepreneur-second characterized many study participants. Across interviews, the joy of making emerged as a main motivator. Participants described deep satisfaction from the tangible nature of their craft. The gem cutter contrasted making with her slower-paced day job: “I can sit down in an evening, finish a gem, hand it off to a satisfied customer... they come back in a few months and show me what they’ve made and it’s very fulfilling to see from the beginning to the end.” This immediacy of feedback and visible impact provided fulfillment beyond financial reward.

The physical nature of making introduced distinctive challenges. Reliance on specific equipment kept some makers tied to particular studio locations. A potter described dependence on external kiln schedules: “If you don’t have your own kiln, you have to go somewhere to ask for firing. But they have their own schedule, so sometimes it takes a really long time.” Material failures added unpredictability: “When the plate is not dry enough, it blows up. And when the fire doesn’t go up enough, the color is weird.” Beyond material failures, makers experienced physical strain. Repetitive motions, awkward postures, and extended hours of focused handwork can lead to chronic pain. Yet for many, this physical engagement is also what makes work meaningful. The study participants described enjoying the embodied experience of manipulating materials, the tactile feedback, the sense of making with one’s own hands. This creates a tension: the same physical engagement that provides satisfaction can, over time, cause harm.

These findings suggest a delicate balance between manual work and technological support. Technology that automates away physical engagement might reduce strain but could simultaneously eliminate the joy that motivates makers. Technology that provides no support leaves makers vulnerable to physical and logistical challenges. This tension motivates the pursuit of mixed-initiative machines, as systems that offer targeted assistance while preserving opportunities for satisfying engagement.

3.1.2 Professional Machine Operators

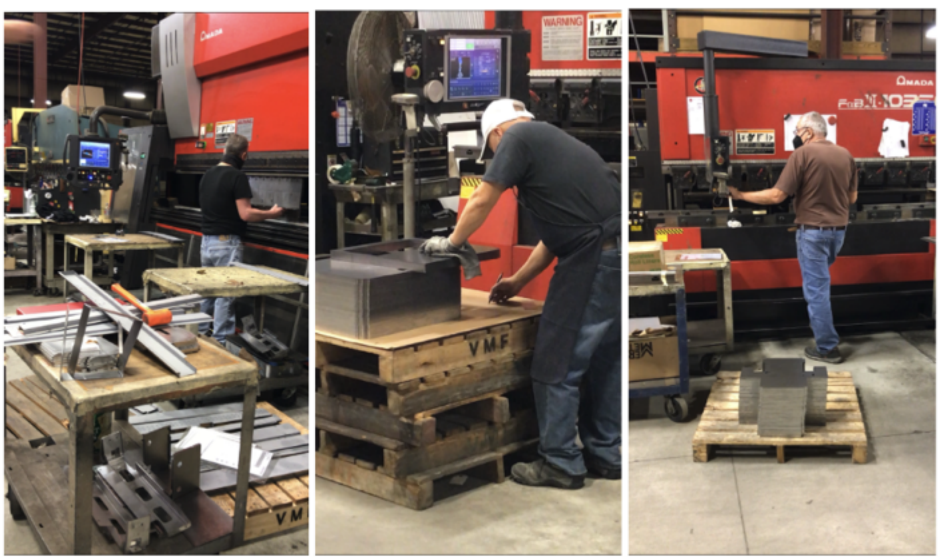


Figure 3.2: Metal workers operating Amada sheet metal bending machines.

To understand how these dynamics play out in professional fabrication contexts, we conducted an observational site visit at a sheet metal bending workshop in Connecticut, where we observed and spoke informally with machine operators across different experience levels. The site had multiple Amada sheet metal bending machines spanning several decades, with interfaces ranging from fully analog, to featuring 3D animations and automatic step tracking.

Workers preferred different machines, different job types, and different workflows. Some preferred smaller, complicated pieces in low volumes; others preferred larger pieces with higher job volumes. One worker above 60 years old, with thirty years of experience, preferred simpler machines and found complex interfaces confusing. He only needed a small subset of the available buttons. He acknowledged that younger workers familiar with smartphones might prefer more complex machines.

A newer worker, three weeks into the job, described still learning the machine. He had made a mistake on a first bend, then repeated the mistake on a new piece because the machine had advanced to the next step. This illustrates how machine state can contribute to errors when operators are still developing familiarity with the task.

The newer worker identified physical strain as a significant challenge. An absence of stools required standing on one leg for eight-hour shifts. He described workers regularly switching which foot operated the machine to manage fatigue. This observation reinforces that physical ergonomics matter also for professional operators, not just hobbyist makers.

The metal workshop observations highlighted individual preferences. The experienced worker wanted simplicity; the newer worker appreciated detailed step displays. This suggests that collaborative fabrication machines should adapt to user expertise and preference rather than providing uniform assistance. The observations also revealed that configuration errors and physical strain are concerns across contexts, not just for novice hobbyists.

3.1.3 Synthesis: What Makers Need

Throughout these interviews and observations, several themes emerged. First of all, makers value the process, not just the product. The joy of making comes from embodied engagement, skill development, and tangible feedback; not merely from having a finished object. Technology that eliminates these experiences may undermine motivation even if it increases efficiency. Second, makers want support that preserves agency. This suggests assistance that catches errors rather than taking over execution. Third, individual preferences differ. Our novice sheet metal worker could welcome guidance that the expert could find intrusive. And finally, both hobbyist makers and professional operators described physical challenges. There may be opportunities for machines to reduce strain in demanding aspects while preserving engagement in satisfying aspects.

While we studied experts to establish real-world needs for mixed-initiative creative support, we want to note that we developed potential interventions with novices. This was partially due to pragmatic considerations: it is easier to recruit novices, and novices function as well in our instrumented workspaces as in any other workspace. Moreover, novices are more similar both in needing help, and the type of help they need. We further revisit this expert-novice gap in the conclusion (see Section 6.2).

3.2 Interaction Vignettes

The interviews and site visit revealed what makers value and what challenges they face, but translating these observations into design requires envisioning concrete interactions. What might it look like for a machine to provide help without taking over? And how could a machine participate in making without disrupting the meditative quality of handwork?

In the next section, we revisit the three quotes that this dissertation opened with. These quotes came from exploratory interviews with hobby makers, and inspired the creation of vignettes to represent various future interactions. These vignettes are conceptual sketches of possible human-machine relationships during making.

3.2.1 Vignette 1: The Machine as Guide

One interviewee described her frustration learning to throw pottery on a wheel. The process itself was enjoyable, but the initial step of centering clay was a persistent obstacle:

> "You know how in bowling there's bumpers? I wish there were that for throwing, to make it so I didn't mess up so much, because I felt like I spent so many hours trying to learn it. [...]"

This suggests a machine that acts as a guide (Figure 3.3), providing support structures that prevent costly errors while preserving the core challenge of skilled work. Imagine drawing a still life of a fruit basket. Rather than worrying about proportions, the plotter provides faint guidelines based on camera observation of the actual objects. The human still draws; the machine ensures the proportions will be right. If the human's later strokes drift from correct proportions, the machine might offer gentle corrective feedback. Guidance should be considered to be selective. The machine handles what the user finds frustrating (centering, proportions) while leaving what the user finds satisfying (the actual throwing, the expressive drawing). The challenge is determining which is which, and recognizing that this varies across users and expertise levels.

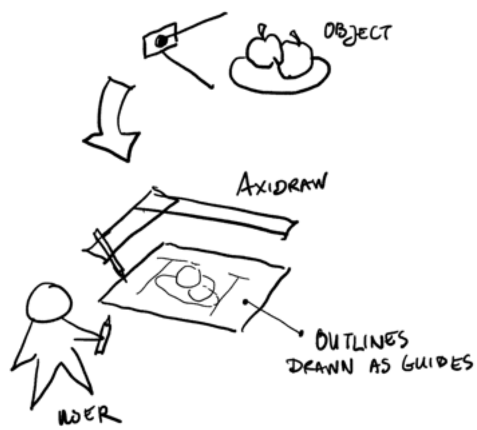


Figure 3.3: The machine as guide: selective initiative that prevents errors while preserving the core challenge of skilled work. The asymmetry of skill is explicit, as the machine provides structure the user does not yet have.

3.2.2 Vignette 2: The Machine as Companion

Another interviewee discussed the meditative quality of handwork:

> "A lot of people, when they feel fidgety, pick up knitting or crochet because it's something they can do with their hands and help them [calm down]."

Here the value lies not in the product but in the process: the rhythmic, absorbing quality of repetitive making. A machine that took over material handling would eliminate precisely what makes the activity worthwhile.

This suggests a machine that acts as a companion (Figure 3.4): a presence that participates without disrupting. Imagine a drawing robot that enhances your doodles through procedurally generated patterns, humming along beside you without demanding attention or taking over. The machine contributes to the emerging artifact but does not direct it. The human remains immersed in the fidgety pleasure of mark-making.

One note here is that machine participation need not mean machine control. A companion contributes without leading. The challenge is designing machine behaviors to feel like companionship rather than disruption or surveillance.

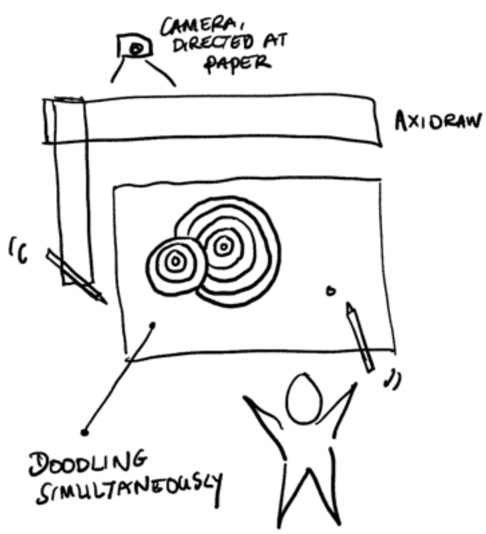


Figure 3.4: The machine as companion: a presence that participates without directing. The machine contributes to the emerging artifact but does not lead; the human remains the primary author.

3.2.3 Vignette 3: The Machine as Adaptive Collaborator

A third interviewee, who worked professionally with 3D printing, expressed frustration with machines not accommodating mid-process intervention:

> "I truly hate not being able to intervene. That's not something I had with clay. I think that many people working in the world of clay now realize that there is value in being able to intervene mid-print. [...] [I see] something I could easily intervene in and save the whole print. But I can't do that because it's not in the code or embedded in the jet. It's not like I can take it out, play with it a little, put it back, and continue."

This suggests a machine that acts as an adaptive collaborator (Figure 3.5): one that observes human interventions and adjusts its behavior accordingly. Imagine a clay printer where both user and machine watch the emerging artifact and each other. When the user reaches in to adjust a sagging section, the machine pauses, observes the modification, and adapts its subsequent tool path. The human and machine are both responding to the material and to each other.

Looking at this, we see that collaboration requires mutual responsiveness. The machine must not only accept human intervention but incorporate it into ongoing behavior. The challenge is sensing what the human has done and reasoning about how to proceed, which requires the perception capabilities that Chapter 4 investigates.

3.2.4 Implications for Design Exploration

These three vignettes (guide, companion, adaptive collaborator) represent different points in the design space of human-machine relationships during making. They differ along several dimensions:

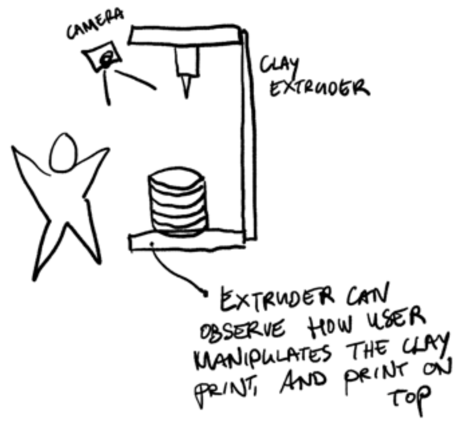


Figure 3.5: The machine as adaptive collaborator: mutual responsiveness between human and machine. When the user intervenes in the material, the machine observes, pauses, and adjusts its subsequent behavior accordingly.

- **Initiative**: The guide takes initiative to provide structure; the companion follows the human's lead; the adaptive collaborator shares initiative with the human dynamically.
- **Attention**: The guide may demand attention (to follow guidelines); the companion minimizes attentional load; the adaptive collaborator requires mutual awareness.
- **Agency**: The guide constrains human action (helpfully); the companion preserves full human agency; the adaptive collaborator negotiates agency moment-to-moment.

The explorations that follow in Section 3.3 do not map one-to-one onto the vignettes; rather, each exploration illuminates different facets of the envisioned relationships. The Tic-Tac-Toe exploration, though adversarial rather than collaborative, demonstrated the social engagement that the companion vignette aspires to: study participants treated the machine as a social presence, not a tool. ShadeBot inverted the guide relationship: rather than the machine guiding the human, the human's line art guided the machine's shading, revealing what it feels like for the machine to follow creative direction. The AxiWOz explorations most directly explored adaptive collaboration, with the wizard simulating the mutual responsiveness the third vignette envisions.

The vignettes also raised questions. The guide vignette asks: can a machine determine which parts of making a user needs help with, and take the lead selectively? Chapter 4 addresses the perceptual half of this question and observes that implicit behavioral cues such as hesitation and postural shifts often precede explicit help-seeking, suggesting that the signals the guide would require are in principle observable. Chapter 5 then tests whether acting on such signals is welcomed.

The companion vignette asks: can a machine participate in making without disrupting its meditative quality? The qualitative findings in Chapter 5 show that study participants valued the machine not only for its technical guidance but for its presence. Several described the machine as providing "company" during an otherwise isolating learning experience.

The adaptive collaborator vignette, which requires the richest perception and the most sophisticated real-time reasoning, remains the furthest from realization. Chapter 4's instrumentation protocol establishes the observational foundation such a system would require, and the CoSew-4 dataset created in Chapter 5 provides data toward the computational models that would eventually enable it. The adaptive collaborator, in this sense, is not a contribution of this dissertation but a direction it points toward.

Together, these vignettes established the design vision that motivated the empirical work. Their purpose was to make the design space visible, which was confirmed, expanded, and in some cases reframed by the studies that followed.

3.3 Design Explorations

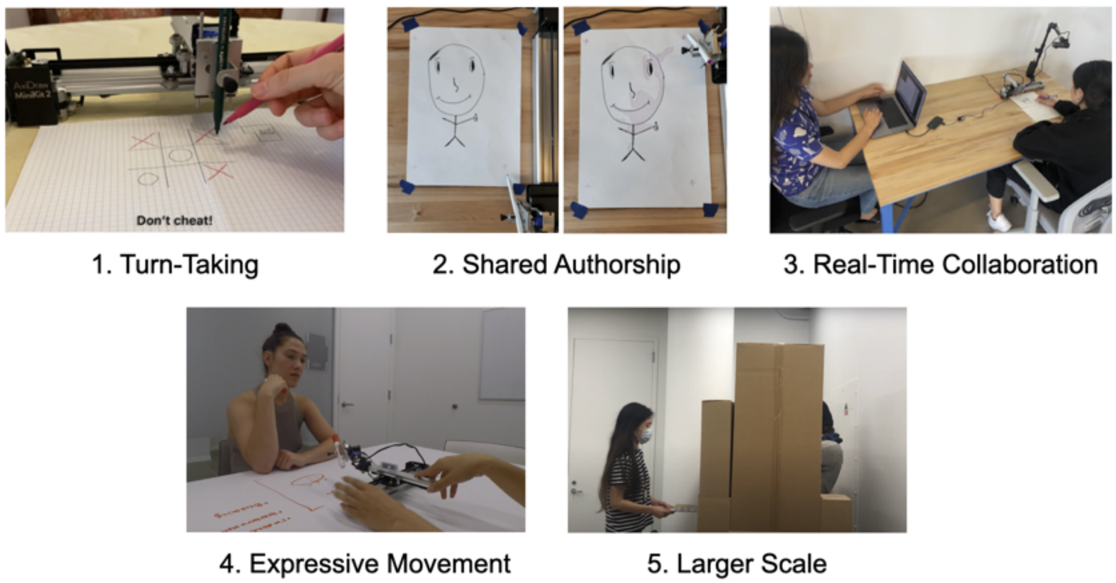


Figure 3.6: We introduce five design explorations on the dimensions of turn-taking, shared authorship, real-time collaboration, expressive movement and larger-scale machines.

Buxton [2010] distinguishes between sketching and prototyping in design. Sketches are quick, tentative, and disposable; their purpose is to explore what to build, not to evaluate how well something works. Prototypes come later, once the design direction is established. The five explorations that follow (see Figure 3.6) are sketches in this sense. Using the AxiDraw pen plotter as a primary platform, supplemented by cardboard mock-ups and Wizard-of-Oz simulations, we conducted a series of rapid design investigations to build intuition about what human-machine collaboration in fabrication could feel like and to surface the dimensions along which such collaboration varies.

Each exploration foregrounded a different aspect of collaboration, informed by the vignettes of Section 3.2. Turn-taking and expressive movement (Explorations 1 and 4) relate to the companion's quality of presence; shared authorship and real-time collaboration (Explorations 2 and 3) engage the guide and adaptive collaborator. Finally, Exploration 5 addresses applications beyond the tabletop by expanding the dimension of scale. This progression was not strictly linear. The explorations were conducted over several semesters and insights flowed between them, but each successive sketch opened questions that the previous ones could not address.

3.3.1 Platform: The AxiDraw Pen Plotter

Most explorations used the AxiDraw pen plotter [Scientist, 2022] as a platform. An example of a typical AxiDraw plot is given in Figure 3.7[2]. The AxiDraw is a precise two-axis machine that moves a pen across paper as well as up and down. It offers advantages for sketching co-creative interactions: it is safe (no cutting, heating, or hazardous materials), relatively fast (enabling rapid iteration), and produces visible artifacts (drawings that can be examined and discussed). The familiarity of pen-on-paper output also reduces barriers to participation.

The AxiDraw's limitations are also instructive. Straight out of the box, it cannot sense, as it has no cameras, pressure sensors, or other input modalities. It cannot respond to what a human collaborator does unless that information is provided externally. These limitations foreground the sensing requirements that Chapter 4 addresses.

[2] The figure was previously published in a doctoral consortium paper as Bremers, A. W. (2022, June). A computer that sketches along with you. In Proceedings of the 14th Conference on Creativity and Cognition (pp. 669-673). [Bremers, 2022a]

Figure 3.7: The result of a design workflow of creating a 3D model in Blender and then plotting it using a rollerball pen. (Image previously published in Bremers [2022a].)

3.3.2 Exploration 1: Turn-Taking with Tic-Tac-Toe

The first exploration simplified collaborative drawing to the game of Tic-Tac-Toe[3]. A human and the AxiDraw took turns marking a shared grid (see Figure 3.8). The system used a webcam and computer vision to recognize the human's marks, then generated and plotted the machine's response.

[3] Avital Dell'Ariccia was a remote research intern whom I supervised in 2021; she carried out the development of Tic-Tac-Toe bot and collected the user study data under my weekly supervision. This work was published in Dell'Ariccia, A., Bremers, A. W., Lee, W. Y., & Ju, W. (2022, February). "Ah! He wants to win!": Social responses to playing Tic-Tac-Toe against a physical drawing robot. In Proceedings of the Sixteenth International Conference on Tangible, Embedded, and Embodied Interaction (pp. 1-6). [Dell'Ariccia et al., 2022b]

We also used the Tic-Tac-Toe system to explore expressive machine movements[4]. Drawing on animation and dance principles (e.g., Takayama et al. [2011]), we developed eleven distinct movement behaviors, such as "confident" moves (direct and fast), "hesitant" moves (slow, wandering), and "frustrated" responses to losing (scribbling chaotically) (see Table B.1). We mapped quantitative movement parameters (velocity, acceleration, path indirectness) to emotional states (confidence, hesitation, frustration, celebration).

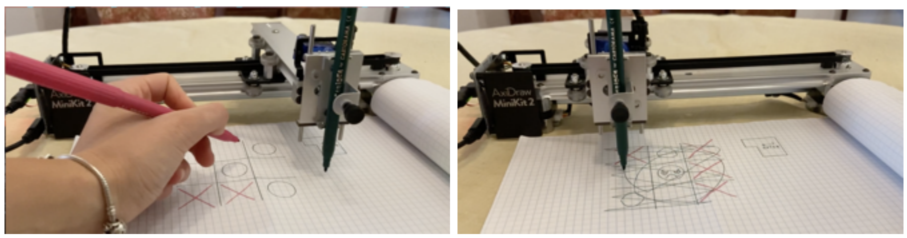


Figure 3.8: A user plays Tic-Tac-Toe against a plotter. Right: The plotter expresses anger when it loses. (Published in [Dell'Ariccia et al., 2022a], © 2022 IEEE.)

3.3.2.1 What We Learned

An exploratory study with three participants showed that all participants verbally addressed the robot during play (e.g., talking to it, exclaiming "Ah! He wants to win!"), displayed careful movements around the robot, and showed signs of competitiveness [Dell'Ariccia et al., 2022b]. The Tic-Tac-Toe exploration confirmed that people readily engage with machines as social actors, consistent with prior work [Nass et al., 1994]. However, the rigid turn-taking structure limited insights about fluid collaboration. Furthermore, as it was a game, we did not expect the user to be invested in the aesthetic result of the task.

[4] This second project from Avital's internship was published in Dell'Ariccia, A., Bremers, A., Michalove, J., & Ju, W. (2022, March). How to Make People Think You're Thinking if You're a Drawing Robot: Expressing Emotions Through the Motions of Writing. In 2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (pp. 1190-1191). IEEE. [Dell'Ariccia et al., 2022a]

3.3.3 Exploration 2: Shared Authorship with ShadeBot

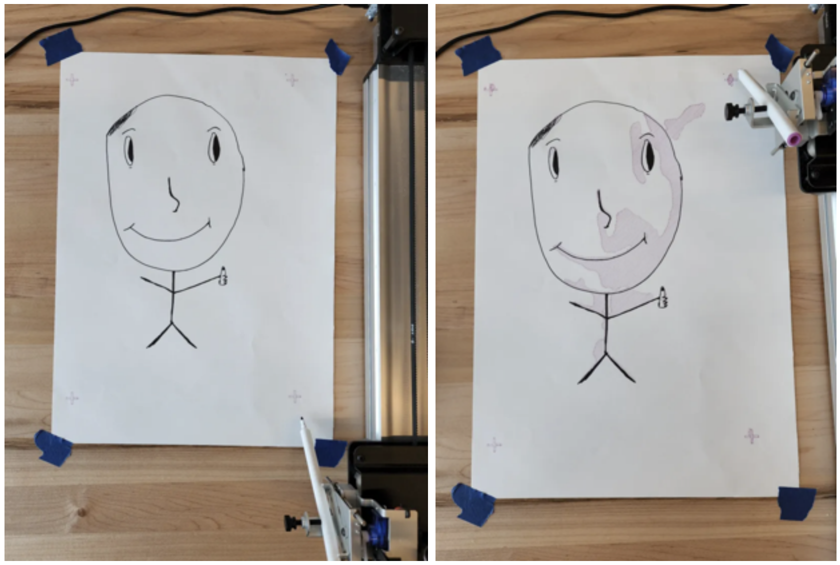


Figure 3.9: ShadeBot explores a physical implementation of a digital shading algorithm, by adding shading to what the user has drawn.

The second exploration moved beyond games to actual drawing. ShadeBot[5] implements a shading algorithm: the human draws line art, and the machine adds shading based on computational analysis (see Figure 3.9). This creates genuine shared authorship: the final drawing reflects both human and machine-originated contributions.

3.3.3.1 What We Learned

Working with ShadeBot revealed the experience of seeing a machine contribute to one's creative work. The shading was not merely executing human intent but adding something the human had not specified. This felt qualitatively different from using a tool.

[5] The technical implementation of ShadeBot was carried out by Cooper Murr, who completed a research internship with me in 2022. I supervised him, set the research direction, and assisted with technical issues.

However, ShadeBot's fixed algorithm limited exploration. The machine always shaded in the same way. There was no room for the machine to adapt to context, respond to human feedback, or vary its approach. This motivated exploring Wizard-of-Oz methods, where a human operator could simulate adaptive machine intelligence.

3.3.4 Exploration 3: Real-Time Collaboration with AxiWOz

Figure 3.10: Using AxiWOz-1, a pilot participant takes turns to draw an apple and a banana. A wizard (out of view, left) uses an iPad to control the plotter.

The third type of exploration used Wizard-of-Oz (see Section 4.2.1 for an explanation of this method) to simulate interactions. We iterated through four different types of wizarding setups, in each of which a wizard observed the interaction and controlled the AxiDraw in real-time. The user believed they were collaborating with an intelligent machine, but the machine's responses were generated by the wizard.

AxiWOz-1 was a system where the wizard would draw on an iPad canvas, and the AxiDraw would follow sections of the same path as it was drawn[6]. This

[6] The development of the iPad interface and the pilot data collection was completed by Tobias Weinberg when he was a research intern. I supervised him, set the research direction, defined system requirements, tested the systems iteratively, and designed the pilot study.

system was piloted with users to draw a still life of two pieces of fruit, depicted in Figure 3.10. While the system did allow for interaction exploration (for example, exposing users to the feeling of sharing paper with a plotter, testing the timing of turn-taking), it suffered from considerable latency, especially when the wizard would draw a quick stroke. One way to reduce latency was to update the pen position less frequently, but this reduced drawing quality, as curves were no longer smooth.

Figure 3.11: A demonstration of AxiWOz-2, which used pre-defined paths to reduce latency. The wizard (left; not hidden in this image) used paths and layers in Inkscape that were iteratively activated.

In order to tackle the latency issue and optimize for better drawing quality, a second approach was tried. For AxiWOz-2, we used pre-designed paths triggered by the wizard using Inkscape paths and layers (Figure 3.11). This method was piloted with users drawing vases, which the machine would fill with flowers, and yielded aesthetically pleasing drawings that were drawn with the same drawing speed as the human, with no noticeable delay.

As a third iteration we built Hero, which included a ZED 2 depth camera for body pose and facial expression tracking, a tabletop-facing webcam, and the AxiDraw, all synchronized through Microsoft's Platform for Situated Intelligence framework[7]. The system was designed with three goals: interaction prototyping (via Wizard-of-Oz), interaction elicitation (capturing the range of behaviors users produce), and data collection (for eventual training of machine learning models). Hero was a precursor to the sensing and data capture methods used in Chapters 4 and 5. Combining the use of the Hero platform with

Figure 3.12: The outcome of a co-creative drawing activity with pre-defined paths using Hero (the "fireworks" were machine-initiated). (Previously published in Bremers and Ju [2024a].)

pre-designed paths, we carried out another interaction pilot. Figure 3.12 shows the resulting drawing when the machine was set up to plot abstract, fan-like shapes to represent flowers. The plotter started plotting the fan-like shapes in the top left area, which was the intended canvas size. The pilot participant was not prompted that these were flowers, and explained after the fact that she interpreted the drawings as fireworks, reminding her of watching fireworks in New

[7] This work was published as Bremers, A., & Ju, W. (2024, October). Can machines tell what people want? Bringing situated intelligence to generative AI. In Proceedings of the Halfway to the Future Symposium (pp. 1-6). [Bremers and Ju, 2024a]

York on the 4th of July and inspiring her to draw a skyline, Statue of Liberty, pizza, and a rat.

Despite the system's limited creative output, we found that open-ended creative interactions were still possible. During the sessions with this system, the pilot participant displayed nonverbal behavior related to drawing stages. Figure 3.13 depicts a sequence of such behavior: before drawing, the user asks for time to brainstorm and practice ideas on a separate sheet of paper. The movements during this ideation process include looking to the side and up and eventually focusing on the paper. This sequence was a precursor to the capturing of user-facing data whilst undergoing proactive assistance that we discuss in Chapter 5.

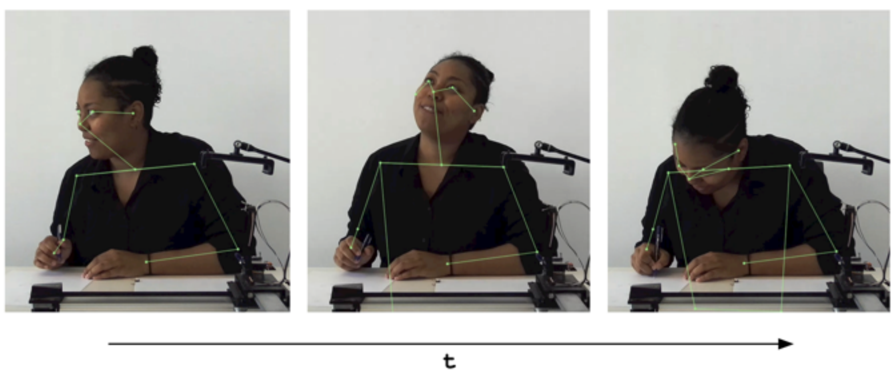


Figure 3.13: Hero, the third iteration of Wizard-of-Oz piloting, allowed us to capture nonverbal interactions of interest, such as a user looking to the side and up while considering what she is going to draw. (Previously published in Bremers and Ju [2024a].)

Finally, the AxiWOz-4 combined pre-designed paths with remote observation. In Figure 3.14, a wizard-of-oz setup is shown where two cameras provided real-time video to the wizard, who was located at a different location from the machine. The AxiDraw in this instance was also designed to have the capability to gesture and wave with the "pen hand" (see Section 3.3.5).

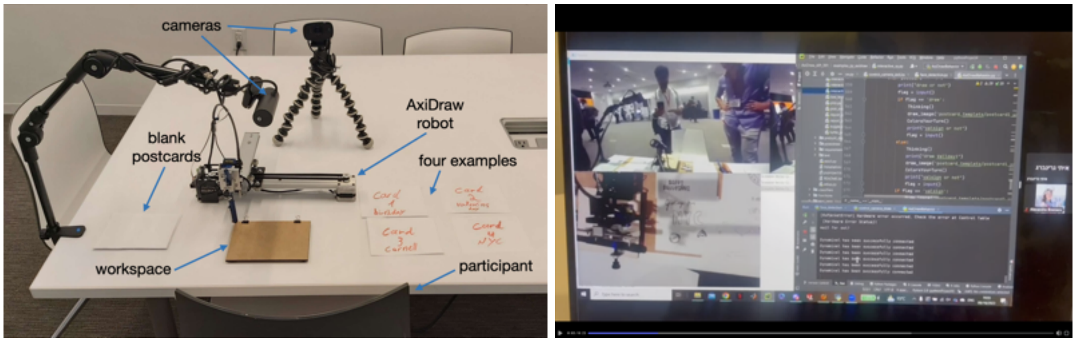


Figure 3.14: In AxiWOz-4, two cameras provided real-time video streams to the wizard (left), who controlled a python program remotely on a computer streaming these views (right). (Partially published in Grinberg et al. [2023].)

3.3.4.1 What We Learned

The Wizard-of-Oz approach enabled exploring interactions beyond what could be technically implemented. The wizard could respond to subtle cues, such as hesitation, confusion, creative choices, that no algorithm could yet detect. This revealed what collaborative fabrication could feel like if machines had sufficient intelligence to act on their own.

The exploration also revealed practical challenges. Where latency in system 1 disrupted the feeling of collaboration, pre-designed paths in system 2 limited responsiveness. These practical challenges highlighted the difficulty of achieving fluid real-time collaboration. In addition to the Tic-Tac-Toe exploration, the AxiWOz explorations also raised the question of the influence of movement and speed on the interaction, which forms the theme of the next section.

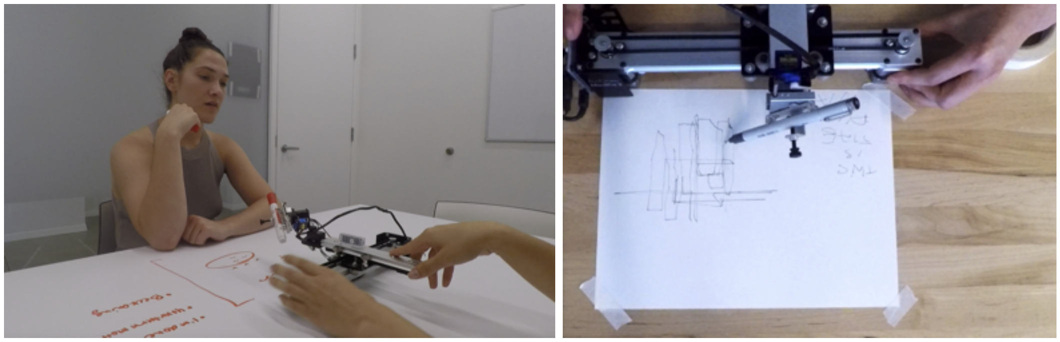


Figure 3.15: Left: a dancer informed the design of movements for a modified AxiDraw with an extra degree of freedom that allowed gestures. Right: an animator puppeteers the machine. (Partially published in Grinberg et al. [2023].)

3.3.5 Exploration 4: Movement Design with AxiDance

Realizing the importance of movement quality from the Tic-Tac-Toe and Axi-WOz explorations, we brought in movement experts. First, we augmented the AxiDraw with two additional servo motors, adding two rotational degrees of freedom to the pen holder[8]. This allowed the machine to gesture (pointing at the user, waving to attract attention, and hovering over the drawing surface to 'look at' the human's work). The design process involved an animator physically puppeteering the modified AxiDraw, and a dancer instructing the interaction designer how to translate her embodied movement ideas into machine commands (see Figure 3.15). This puppeteering method yielded a structured interaction protocol organized into two phases: a welcoming phase (where the machine uses noise, pointing, and gaze-following to engage the user) and a collaborative drawing phase (where the machine alternates between plotting and emotive gestures, observes the human's drawing by hovering, and signals turn

[8] The extra degrees of freedom were added to the plotter by Itay Grinberg, while he was a research intern under my supervision. This work led to the publication Grinberg, I., Bremers, A., Pancoast, L., & Ju, W. (2023, October). Implicit collaboration with a drawing machine through dance movements. In Proceedings of the 8th ACM Symposium on Computational Fabrication (pp. 1-2). [Grinberg et al., 2023]

boundaries). The protocol, documented in Tables C.1 and C.2, represents a concrete design artifact for machine social behavior during co-creative activity.

3.3.5.1 What We Learned

The dancers and animators generated movement vocabularies that engineers would not have conceived. A dancer demonstrated how the machine might “look at” the human’s drawing by hovering over it before proceeding. An animator showed how anticipatory movements could signal upcoming actions.

All explorations to this point used the tabletop AxiDraw. The final exploration asked whether insights about collaboration, movement, and timing would hold at the scale of industrial fabrication machines.

3.3.6 Exploration 5: Sketching Scale Using Cardboard

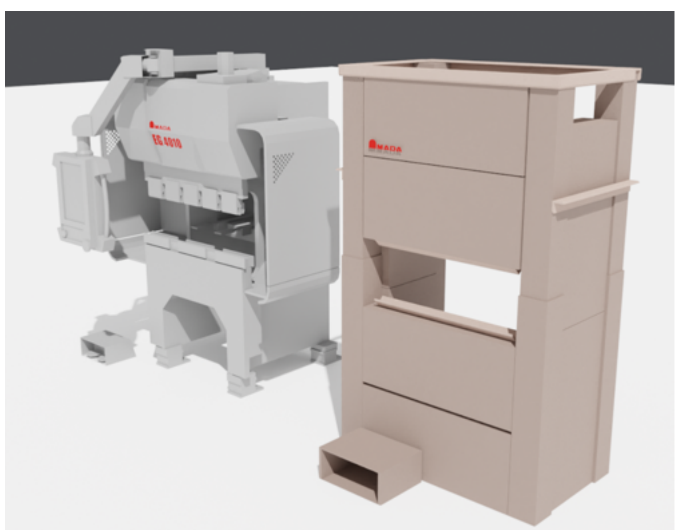


Figure 3.16: A 3D model was used to design a cardboard scale replica of an Amada sheet metal bending machine.

The final exploration investigated larger-scale fabrication machines by the use of cardboard [Peek et al., 2017]. Using a life-size cardboard mock-up of a sheet metal press brake (depicted in Figure 3.16[9]), informed by metal workshop observations, we explored various ways to interact. We conducted pilot tests where a wizard behind the cardboard operated the "machine" while study participants performed bending tasks (demonstrated in Figure 3.17).[10]

Figure 3.17: A user stands in front of the machine with a piece of cardboard that is to be bent. A wizard (right) stands behind the machine and lowers the top half in order to simulate the bending operation of a sheet metal press brake.

3.3.6.1 What We Learned

The cardboard model revealed how scale and embodiment affect interaction. Standing at a large machine feels different from sitting at a tabletop plotter. The physical presence of the machine (e.g., its size, its sounds, its movements) shapes the experience in ways that cannot be captured in smaller-scale models.

[9] The 3D model and physical build were completed by my collaborator David Goedicke.

[10] Pilot test data collection was carried out by Daniel Zhou while he was an intern. I supervised him, and designed the task and data collection protocol.

The exploration also demonstrated that low-fidelity mock-ups can elicit meaningful interaction data. Study participants engaged seriously with the cardboard machine, enabling us to observe how they would approach assistance from a large fabrication system.

3.3.7 Synthesis: Lessons from Design Exploration

Across these design sketches, five dimensions of the design space for collaborative fabrication became visible. First, movement quality matters independently of outcome. The Tic-Tac-Toe and AxiDance explorations demonstrated that expressive movements communicate intention and engagement. A "hesitant" approach to the drawing surface feels different from a "confident" one, even when the mark produced is identical. Second, shared authorship changes the interaction. The ShadeBot exploration revealed the qualitative difference between using a machine as a tool and experiencing it as a contributor. Third, sensing is a prerequisite for initiative. Out of the box, the AxiDraw cannot take meaningful initiative because it cannot sense. The AxiWOz explorations demonstrated what collaboration could feel like with sufficient sensing capability, while simultaneously exposing how much the wizard relied on real-time observation of the study participant's behavior, attention, and affect. Fourth, timing and appropriateness are challenges. Even with wizard control, determining when to intervene and how proved difficult. Interventions that arrived a few seconds too late lost their value; interventions that arrived unprompted sometimes disrupted flow. Fifth, the cardboard exploration demonstrated that physical scale changes the interaction.

These explorations varied in their directness of influence on subsequent chapters. The AxiWOz iterations (Section 3.3.4) most directly informed the instrumentation protocol of Chapter 4 and the Wizard-of-Oz study design of Chapter 5. Tic-Tac-Toe (Section 3.3.2) and ShadeBot (Section 3.3.3) provided foundational insights about social engagement and shared authorship. Movement explorations (Section 3.3.5) and the cardboard mock-up (Section 3.3.6) established dimensions of expressive movement and physical scale, which were not further explored in Chapter 4 and Chapter 5, but remain open for future empirical investigation.

3.4 Summary: The Design Space and Open Questions

This chapter has investigated the design space for mixed-initiative fabrication machines through interviews, observations, conceptual vignettes, and a series of design sketches. Empirical work with makers confirms that they want support that preserves agency. Design explorations illustrate the space of possibilities and identify challenges. Of the five dimensions listed in Section 3.3.7, two (sensing and timing) are the focus of the next chapters:

1. How can researchers capture the interactions around machines? The design explorations revealed that machines need to perceive user states to take appropriate initiative. Chapter 4 addresses this question through an investigation of multi-camera instrumentation for tabletop fabrication.
2. Is proactive assistance beneficial? The theoretical analysis suggests that well-timed machine initiative could support makers, but that users might experience proactive assistance as intrusive. Chapter 5 addresses this

question through a comparison of proactive and reactive assistance with novice users.

The remaining three dimensions (movement quality, shared authorship, and physical scale) are outside the main scope of this dissertation and point toward future work.

CHAPTER 4

PART II: CAPTURING INTERACTIONS AROUND MACHINES

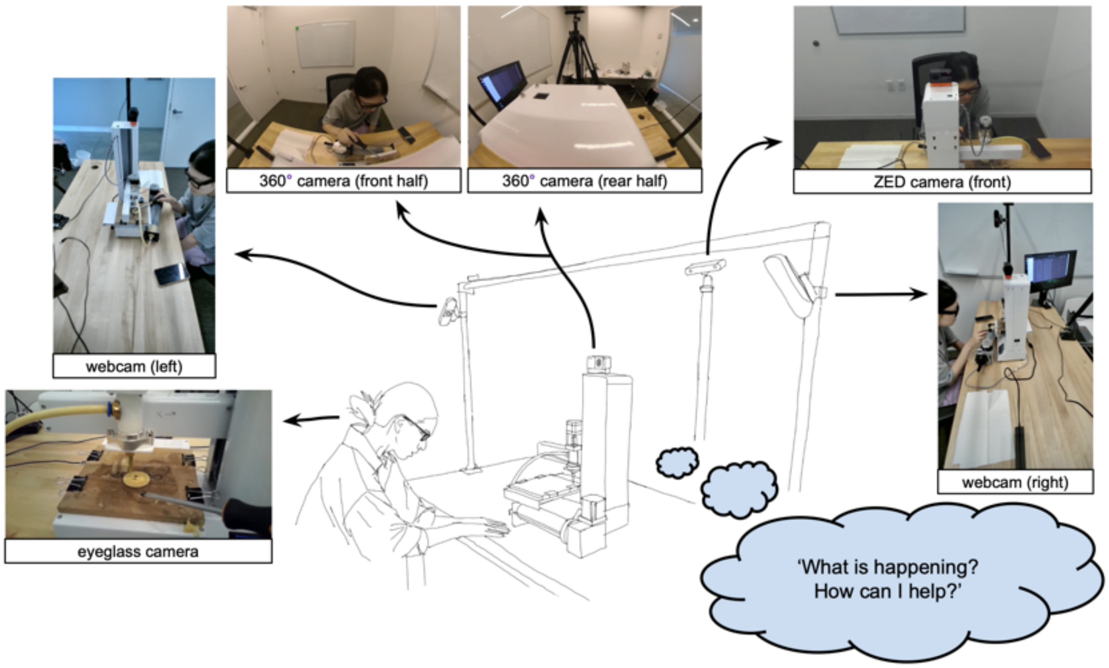


Figure 4.1: We analyzed if and how cameras can capture interactions with fabrication machines. Depicted are all camera views around a study participant interacting with a clay extruder.

Designing effective interactions for co-creative machines requires an understanding of how people engage with these machines, which starts with eliciting and capturing interactions. In most cases of interaction design, relevant information of interest includes explicit signals (e.g., verbal utterances) and implicit signals (e.g., expressive movements) of the user. Physically co-creative fabrication machines introduce dimensions of physicality, such as material properties and human- and machine-specific constraints, into co-creative interactions. This opens up novel opportunities for interaction design, taking into account user state, machine state (e.g., settings [Subbaraman et al., 2025]) and artifact state. It also raises a methodological question: how can researchers observe and capture these interactions without disrupting natural making behavior?

In this chapter, we[1] investigate (1) whether multi-view camera instrumentation can serve as an accessible and systematic method for capturing human-machine interactions during fabrication; and (2) what these observations reveal about the signals and situations that characterize fabrication interactions, particularly those that might inform the design of initiative-taking machines. We recorded study participants (*N*=12) interacting with each of three fabrication machines: a clay extruder, a pen plotter, and a sewing machine. For each machine, we discuss the potential actions, interaction affordances, and indicators of opportune moments for machine interventions (e.g., indications of error).

From our synchronized video and audio streams from the perspectives of the room (monoscopic video), the person (camera eyeglasses), and the machine (360° video) (Figure 4.3), we find that machine-based user-oriented video provides a clear view of the user's facial expression and body pose with minimal occlusion, while egocentric video shows machine control and material adjustment as most novices look where they are making adjustments. Room cameras can add insights into machine states when the field of view or occlusion from the egocentric and machine-based cameras is not ideal, such as around the handwheel of a sewing machine, or with more expert user groups whose gaze might not always follow their hands' actions. Recognizing that ideal instrumentation will depend on task and context, we propose a protocol and method that interaction designers can use to determine the right instrumentation for capturing interactions around tabletop-scale fabrication machines. We also share our recommendations for various camera setups depending on research aims.

This study focuses on novice users interacting with tabletop-scale fabrica-

[1] This work is intended for future submission to publication, likely as Bremers, A., Roumen, T., Guimbretière, F., & Ju, W. (2026). Instrumenting Fabrication Machines for Video Interaction Analysis and Creative Support. Venue to be determined.

tion machines, captured primarily through video instrumentation. We acknowledge that other sensing modalities (touch, audio, physiological sensors, machine logs), other skill levels, and other machine scales would extend the space of possible instrumentation approaches. Our aim is not to map the full space of fabrication instrumentation but to develop and validate a camera-based approach that is practical, replicable, and sufficient for the interaction analysis needs of researchers studying co-creative fabrication.

Our contributions include: (1) an evaluation of multi-camera instrumentation for capturing human-machine fabrication interactions, demonstrating that a minimal combination of egocentric and machine-mounted cameras provides sufficient observational coverage; (2) a generalizable protocol for instrumenting fabrication machines for interaction analysis; (3) FabriCam-5, a 5-camera synchronized multi-view dataset of novice interactions with fabrication machines (to be published on Harvard Dataverse: [Bremers et al., 2025]), and (4) initial analysis of the FabriCam-5 dataset to uncover interaction signals and situations as they apply to each of the machines.

4.1 Implicit and Explicit Human-Machine Interaction

Building on Ju's framework for implicit interaction design (Chapter 2), where systems show initiative while requiring lower attentional demand, this section considers how implicit and explicit signals manifest specifically in fabrication contexts, and what that means for instrumentation.

Prior efforts have looked into using the expressive qualities of non-anthropomorphic robot movement to communicate and collaborate with people [Takayama et al., 2011, Hoffman and Ju, 2014, Mok et al., 2015, Li et al., 2019, Bethel and Murphy, 2007]. There is a myriad of social cues that people use with each other to communicate context, failure, and success, ranging from subtle cues like eye gaze to more overt cues like language and gestures [Wang et al., 2022, Lin et al., 2020a]. People draw from years of interactive experiences to detect, interpret, and act on these cues from other humans. Researchers have attempted to simulate this sensitivity in technological interactions. However, the ability of robots to understand implicit information and social cues remains underdeveloped as of recent surveys [Sandini and Sciutti, 2018]. Prior work on conversational interactions found that human facial reactions to voice assistant errors can help autonomous systems determine when an error has been made [Cuadra et al., 2021] and that nonverbal reactions to conversational errors can be used to detect dialogue breakdown [Kontogiorgos et al., 2021]. While verbal responses are powerful social cues, non-verbal behavior is essential to human communication [Burgoon et al., 1984].

RemoteCode [Sakashita et al., 2022] is an example of a recent system that brings implicit, non-verbal information to a computer-mediated collaborative interaction. RemoteCode consists of a telepresence system that allows remote collaborators, who would normally only see each other's faces on a static video display, to see each other's head movements directly translated to the head movements of the telepresence robot. In this way, users can see whether the remote collaborator is paying attention to their person, or to the task at hand.

4.1.1 Interactive Aspects of Co-Creative Machines

As discussed in Chapter 2, co-creative fabrication machines go beyond interactive fabrication by striving toward mixed-initiative interaction. Designing such machines requires attention to interaction quality factors including legibility [Dragan et al., 2013, Srinivasa et al., 2017], initiative [Ju and Leifer, 2008, Horvitz, 1999], attentional demand [Olsen and Goodrich, 2003], and timing [Semmens et al., 2019].

Tabletop-scale fabrication machines such as clay extruders, pen plotters, and sewing machines (Figure 4.4) combine digital control and embodied material practice. These machines combine precise computational actuation with continuous physical processes that unfold in real time. In contrast to screen-based interfaces, these machines operate under physical constraints and typically require human attention to jointly shape an evolving artifact. Several aspects make these fabrication machines particularly challenging for interaction analysis. First, many machines require fine-grained manual control of settings (e.g., feed rate), either through buttons on a digital interface or via knobs (e.g., thread tension). Second, the tasks are constrained by material properties. Depending on the material, tasks cannot easily be undone, which makes timing and error prevention crucial aspects of interaction. Third, fabrication tasks often involve both creative fabrication and tacit knowledge, making it hard to define task success beforehand, as expressivity or aesthetic qualities might carry just as much or even more importance than efficiency and mechanical precision.

4.2 Methods of Studying Human-Machine Interactions

This section reviews the methodological approaches of Wizard-of-Oz for eliciting responses to systems not yet built, and video analysis for examining interaction data.

4.2.1 Wizard-of-Oz

To elicit user responses to systems that have not been designed yet, human-computer interaction (HCI) and human-robot interaction (HRI) researchers use Wizard-of-Oz (WOz) methods, employing a researcher-in-the-loop to simulate system capabilities [Dahlbäck et al., 1993]. WOz is widely used as a method of interaction elicitation, allowing designers to test out interactions with users before they are fully developed and implemented (e.g., Martelaro and Ju [2017], Sirkin et al. [2015], Zamfirescu-Pereira et al. [2021]). Typically, a WOz study consists of an interaction performed by a hidden experimenter ("the wizard") to elicit user responses to it. This approach was first used in natural-language interaction to elicit a data corpus of what people would say to a natural-language agent [Dahlbäck et al., 1993]. Horvitz et al. [2013] used WOz techniques to elicit the types of problems people might have using Excel and Word, and to learn what kinds of assistance might be helpful from an interactive agent. In HRI, WOz is most commonly used for simulating verbal (72.2%) and non-verbal (48.1%) processing [Riek, 2012]; our use here differs in being focused on generating expressive motion and eliciting user action.

For fabrication machines, a challenge identified through prior design work is that robots must manage two responsibilities simultaneously: performing the creative task, and collaborating with the person in a way that supports rather than disrupts their ability to do the task [Blackwell, 2015, Amershi et al., 2019]. With physical creative tasks, this means negotiating shared resources such as influence, space, and order of interaction in situations where there may be no clear turns, endpoints, or starting points. These open-ended dynamics make Wizard-of-Oz a particularly suitable research method [Dahlbäck et al., 1993], allowing researchers to simulate machine behaviors before they are technically implemented. In WOz experiments, the setup is usually instrumented with cameras that capture the interaction as it unfolds, so that video analysis [Jordan and Henderson, 1995] can occur at a later moment. WOz studies often require

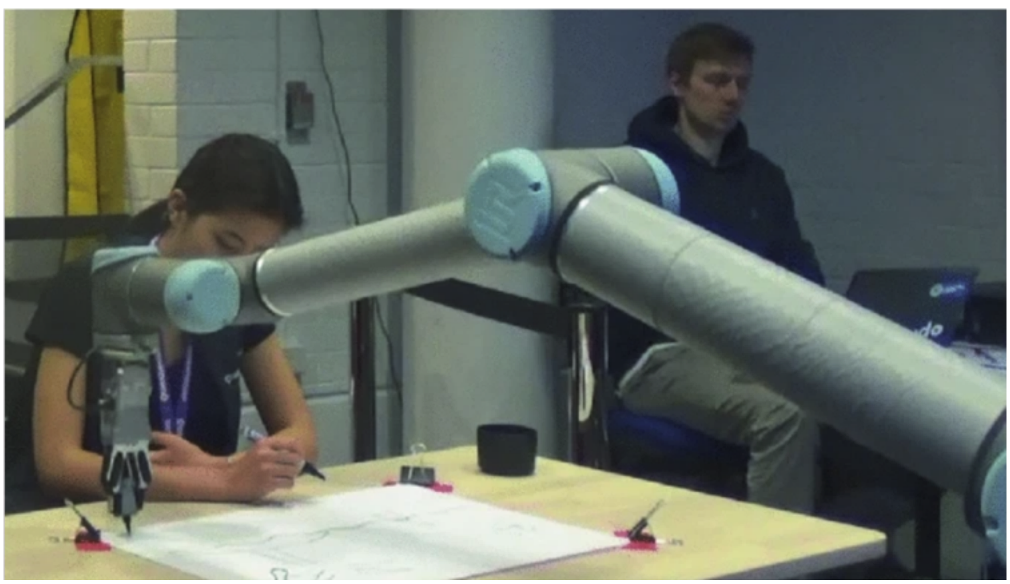


Figure 4.2: A wizard-of-oz approach for collaborative drawing, where the wizard sits in a corner controlling a robotic arm. (Image reprinted from Hinwood et al. [2018], © 2018 Springer Nature Switzerland AG.)

careful planning [Porcheron et al., 2021] and are typically tailored to specific machines. For instance, Hinwood et al. [2018] employed a WOz setup with a UR-10 robot arm to study collaborative drawing, using bespoke instrumentation to prototype and analyze interactions (see Figure 4.2). Fabrication machines have detailed controls and can be difficult to instrument in a way that captures all the required information to analyze interactions, without introducing occlusions that obscure relevant views. Furthermore, tasks are often open-ended and highly variable. After the human-system encounter, interviews [Interaction Design Foundation, 2025] can be used to gather additional qualitative insights into user needs. The captured data can be analyzed post-hoc using interaction analysis methods, which can include video analysis [Suchman and Trigg, 2020, Jordan and Henderson, 1995], linkographic analysis [Goldschmidt, 2014], and artifact analysis [Trausan-Matu and Slotta, 2021]. With the rapid development of computational methods of data analysis, interaction data is increasingly valuable as a way to not only understand, but also model and predict human-system interactions.

4.2.2 Video Analysis

Designers can rely on a variety of methods to gather data for interaction analysis, both within the needfinding step of the design process and in later stages. Human-computer interaction researchers have used ethnographic methods and methods from ethnomethodology and conversation analysis (EMCA) to understand the social experiences and practices of interacting with a certain technology (e.g., the experience of being in a car [Laurier et al., 2008]). Findings from ethnographies, together with interaction excerpts and analyses, provide designers with detailed insights to guide their work.

Using interaction analysis, as specified by Jordan and Henderson [1995], groups of researchers extract richer insights from the observed interactions.

Usually, video analytic approaches focus on detailed analysis of action sequences, for instance, as seen in Laurier et al. [2008, 2012]. This video capture and careful analysis can help researchers develop an intricate and grounded understanding of how these interactions unfold, providing rich insights for system design [Mackay et al., 1988]. Although video remains a strong and frequently used medium for the analysis of interactions, as well as in ethnomethodological research, video fundamentally is a transformation of reality that can never capture all its fullness [Jordan and Henderson, 1995].

4.3 Instrumenting Machines for Interaction Design

Instrumentation of interactive systems is essential for HCI, HRI and design research, in order to systematically capture and analyze interactions between people and computational artifacts. By instrumentation, we here refer to the hardware and software that is used to record user behavior (e.g., gestures, utterances), system states (e.g., machine parameters and error logs), and contextual information (e.g., environmental information and task status).

A body of work in interaction design emphasizes audiovisual data captured by one or more cameras. To mitigate issues like occlusion, researchers can incorporate mobile or egocentric cameras (e.g., Glăveanu and Lahlou [2012]), 360° cameras (e.g., Rajarathinam et al. [2024], Brown et al. [2024]) or artifact-based cameras (e.g., Brown et al. [2024]). Data captured can be analyzed manually (e.g., following ethnomethodological principles [Pelikan, 2023] and video analysis methods [Jordan and Henderson, 1995, Suchman, 1987]), for example applied through studies on trashbarrel robots [Bu et al., 2025], robot sound [Pelikan and Jung, 2023] and robot emotion [Pelikan et al., 2020], or through computational approaches (including VLM-augmented storyboarding by Bu and Ju [2025] or video annotation through object detection [Bianco et al., 2015]). Fabrication activities present particular challenges to interaction capture, such as the need to capture detailed hand movements, material transformations and rapidly changing machine states. Using combined camera perspectives these issues could potentially be mitigated: egocentric cameras reveal user attention and manipulation; machine-mounted cameras show the workspace from the machine's vantage point; and contextual room cameras capture whole-body movements and spatial coordination.

Beyond audiovisual data, researchers can complement their data with additional quantitative data streams that capture traces that are otherwise invisible to cameras. Examples of quantitative data streams can come from physiological sensors, and machine settings and logs (e.g., provided through an API [Evil Mad Scientist, 2025] or from embedded sensors [Clapp et al., 1992]). A combination of quantitative and qualitative data can enable mixed-methods analyses which link subjective observations with objective measures of error or performance, which could in turn facilitate the development of machine learning models (e.g., Bremers et al. [2023b], Semmens et al. [2019], Caber et al. [2023]). However, picking the right instrumentation involves practical trade-offs to balance completeness with facilitating the natural flow of interaction, as well as ease and reliability of operation by the experimenter. Specifically for fabrication machines, an API is not always available; people might use machines they have owned for a long time, and the research interest might lie in understanding existing machine use rather than studying the newest types with advanced logging features. Our study here contributes to the interaction design methodology literature by exploring audiovisual instrumentation suitable for analyzing co-creative human-machine interactions.

4.3.1 Design Requirements

The design of our instrumentation system was guided by requirements based on the interaction analysis literature, as well as the practical constraints of studying fabrication tasks.

First, the system needed to enable comprehensive signal capture. Analyzing human-machine interactions requires access to human state signals (facial expressions, body pose, gaze direction, gestures), machine state signals (operation modes, settings, actions), artifact state signals (material transformations, emerging forms), and contextual information (spatial relationships, tool use). Single-camera setups risk missing signals due to occlusion or limited field of view, particularly when study participants move or when machines have complex geometries.

Second, the system required temporal synchronization. Fabrication interactions unfold dynamically, with human reactions often occurring quickly following machine actions or material failures. To facilitate fine-grained analysis, all data streams needed to share a common temporal reference.

Third, the instrumentation should maintain minimal intrusiveness. Cameras and sensors should capture interactions without creating physical barriers, generating distracting noise, or otherwise altering how users engage with the machines. This is particularly important for creative tasks, where feelings of immersion and flow are large contributors to positive user experiences.

Fourth, the system needed practical operability. The setup should be reliable across multiple sessions and manageable by a single experimenter. This requirement led us to balance comprehensiveness against complexity, avoiding configurations requiring constant troubleshooting or multiple operators.

Finally, the instrumentation needed to support flexibility across machines. Since we studied three fabrication machines with different form factors and interaction patterns, the system needed to accommodate varying workspace configurations while maintaining consistent data quality. Camera positions and angles needed to be adjustable without requiring complete reconfiguration between tasks.

These requirements informed the specific hardware choices, mounting positions, and data capture workflow described in the following section.

4.3.2 Instrumentation Design

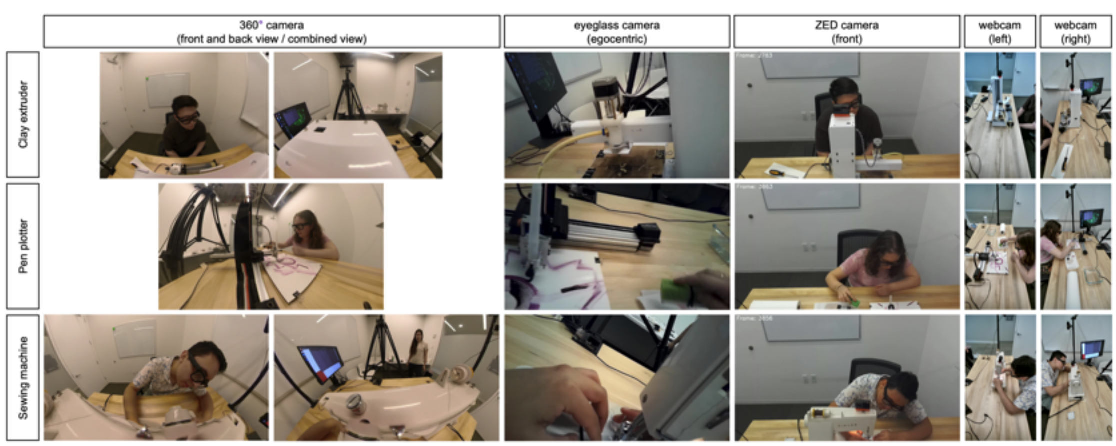


Figure 4.3: 360° video was exported to a front and back view, or a combined wide angle view for the pen plotter. Each row shows the extracted frame at one specific point in time for all 5 cameras (+- 1 second variation).

Our platform centered on an NVIDIA Jetson Orin Nano, with an externally powered USB hub, running the Platform for Situated Intelligence (PSI) [Microsoft, 2024] to coordinate multiple Python scripts for gathering data streams[2]. Two 2K webcams were mounted overhead on a stable rig to capture left- and right-side video and audio; a ZED 2 camera on a tripod recorded user-facing monoscopic video. To capture a 360° machine viewpoint, we Velcro-mounted an Insta360 camera on each fabrication device. Study participants also wore pinhole camera eyeglasses[3] for a user-perspective view. The webcam and ZED 2 video streams were stored on the Jetson Orin Nano, which was controlled via a connected monitor, keyboard, and mouse. Pinhole camera and 360° video were stored on-device. We set up the system in a room at Cornell Tech, adjusting each camera angle to ensure complete coverage of the participant. A schematic view of the setup and all camera angles is shown in Figure 4.1, with specific views for each machine depicted in Figure 4.3

4.4 User Study

To understand how to facilitate designing collaborative interactions with fabrication machines, we conducted a mixed-methods user study. We captured FabriCam-5: a multi-modal dataset of 12 study participants interacting with three different machines, yielding three hours and 48 minutes of synchronized data for each of five camera streams. The dataset includes static monoscopic video from three angles (upper left, upper right, front), participant-perspective footage from pinhole camera eyeglasses, and 360° video from the machine's viewpoint. Audio was recorded by one static camera, the eyeglasses, and the

[2] PSI only handles logging, but it could be expanded to include machine-behavior tracking.
[3] We used Zetronix Kestrel Pro 128 GB - 1080p WIFI HD Video Camera Sunglasses.

360° camera. In addition, we collected interview data to evaluate participants' subjective experiences with the tasks and instrumentation. The protocol was reviewed by Cornell University IRB under exempt number #IRB0148868.

Study participants performed tasks on three fabrication machines: modeling a cylinder using a clay extruder, painting with watercolor using a pen plotter, and sewing together two pieces of fabric using a sewing machine (see Figure 4.5). On each machine, the outcome of the creative work depended both on performance of the machine (i.e., a rigid program), and on how the participant molded physical materials to adapt to the behavior of the machine. We recorded the users' interaction with the machine through a setup of various cameras, to be used in later interaction analysis (see Figure 4.1). After the tasks, we conducted a semi-structured qualitative interview (questions in Appendix D), in order to determine the aspects of the interaction with the machines that participants subjectively deemed to be the most relevant. The interviews were audio-recorded and transcribed using Whisper [OpenAI, 2022] on a local GPU.

4.4.1 Setup

The study was conducted in a laboratory space at Cornell Tech, arranged with a desk positioned for study participants to sit comfortably while interacting with each machine. Desk height and seating were the same across all sessions. Each fabrication machine was placed sequentially on the desk during its corresponding task, with supplementary materials (tools, materials) positioned within reach. Lighting was kept consistent throughout sessions using overhead fixtures in the room.

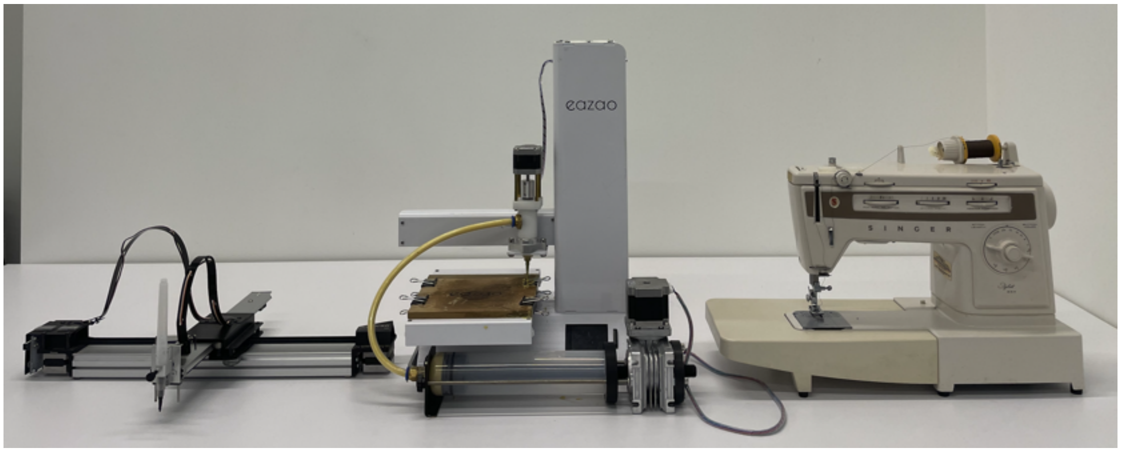


Figure 4.4: We selected three tabletop-based fabrication machines: a pen plotter (AxiDraw V3), a clay extruder (Eazao Zero) and a sewing machine (Singer Stylist 834).

The camera rig was positioned to capture participants from multiple angles without obstructing their access to the machines or materials. Two 2K webcams were mounted on an adjustable aluminum frame spanning the workspace, angled to capture the left and right sides of the interaction area. The ZED 2 camera was placed on a tripod approximately one meter in front of the participant's seated position, oriented to capture a frontal view. The Insta360 camera was repositioned for each machine, Velcro-mounted to ensure a participant-facing perspective from the machine's vantage point. Participants wore the pinhole camera eyeglasses throughout all three tasks.

The experimenter's station was located near the floor opposite of the participant, providing access to the laptop controlling the AxiDraw and the Jetson Orin Nano recording system. This position allowed the experimenter to monitor recording status and control the pen plotter task while remaining outside the primary camera views. The experimenter intervened only when necessary for machine control (pen plotter) or participant safety (sewing machine).

The physical arrangement is depicted schematically in Figure 4.1 and specific camera views for each machine are shown in Figure 4.3.

4.4.2 Task Design

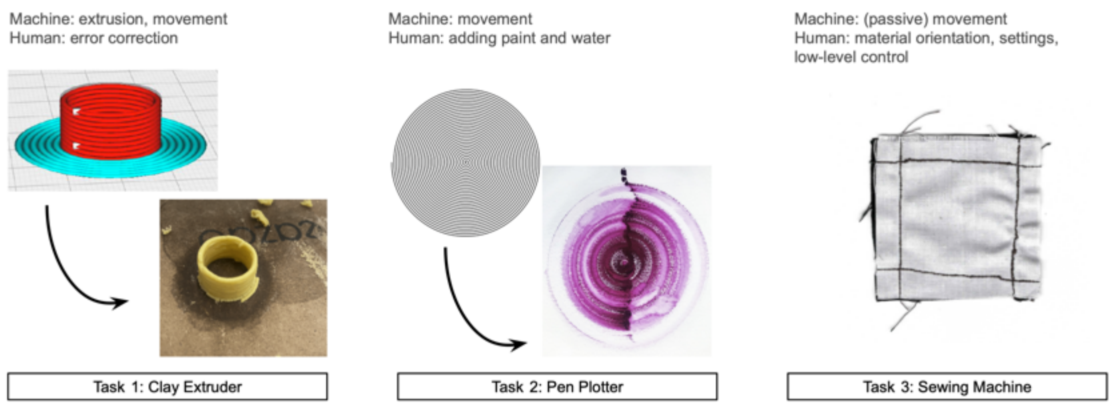


Figure 4.5: We designed three tasks for human-machine collaboration. All tasks took place on tabletop-based fabrication machines and were estimated to take up to 10 minutes each.

We selected three commercially available, tabletop-based fabrication machines (Figure 4.4). For each machine we designed a maximum 10-minute task, to balance low barriers to entry, varying artifact complexity (one, two, or three degrees of freedom), and differing familiarity (sewing machines are common, while clay extruders are rarer). To keep operation straightforward, we simplified the geometry for each machine: the Eazao extruded a cylinder (a goal-oriented task intentionally designed with a high risk of failure despite human intervention), the AxiDraw plotted a spiral (an open-ended task, leveraging participants' experiences with painting but relying on participants to control paint and water), and the sewing machine performed a simple straight stitch on (uniform and non-stretch) quilting cotton, with success largely depending on and defined by the user. Tasks are visualized in Figure 4.5 detailed below.

Using the Eazao Zero clay extruder, the study participants attempted to create a cylinder from wheat-and-corn dough (chosen for dust sensitivity constraints, cf. Buechley and Ta [2023]). The cylinder pattern was sliced in Ultimaker Cura and controlled by the experimenter via the machine's touchscreen. Participants received a screwdriver, paper towels, and could use their hands. The shape was designed to have a high probability of failure in order to elicit participant interactions. Due to dough inconsistencies and the additive nature of the task, it was common for the substrate to be either too dry, resulting in slowed extrusion and lack of layer adherence, or too liquid, which resulted in structural instability of the end print. Due to these characteristics, we expected participants to react to failures through communication or attempted mitigation. In addition, we expected the task to prompt frustration, stress, and potentially anticipatory signals due to the slow nature and build-up of failed prints.

Using an AxiDraw V3 pen plotter, we introduced watercolor paint to add unpredictable behavior and minimize skill gaps between users and the precisely controlled machine. Instead of a pen, the plotter held a vertically oriented synthetic-bristle brush (chosen for durability under repeated forces), kept empty of its built-in water reserve. We designed an outwardly spiraling path – predictable enough for participants to anticipate its motion yet abstract enough to enable creative choices. Participants received a tube of watercolor paint[4], a bowl of water, and paper towels. The paper was attached to the table, and a sponge on the brush occasionally dripped water, further increasing unpredictability. The experimenter controlled the machine via Inkscape's AxiDraw Control extension on a laptop connected over USB. We expected this watercolor-based task to be enjoyable, low-pressure, and conducive to improvisation.

[4] Daniel Smith Rose of Ultramarine was chosen for its two pigments that react differently to varying water amounts, thus adding more unpredictability to the interaction.

Unlike the digitally controlled extruder and pen plotter, the Singer Stylist 834 (circa 1987) is a fully analog sewing machine with an electric foot pedal and no digital interface. Although sewing machines are arguably the most common household fabrication device, they can still fail in unpredictable ways (e.g., thread knots, broken needles, tension issues). Ideally, the machine only moves its needle up and down at a set speed, but mechanical issues often occur. In this task, participants were asked to attach two 2.5-inch squares of quilting cotton (white side up). They were shown an example using four ¼-inch seams but could decide their own sewing path to attach the pieces of fabric. Participants unfamiliar with sewing received a brief orientation, and everyone was warned to keep fingers away from the needle area.

4.4.3 Procedure

The three tasks (clay extruder, pen plotter, sewing machine) were always completed in this order. Before starting, the setup was arranged with the first machine (clay extruder) on the table, and the experimenter verified the Jetson Orin Nano's readiness (this involved checking free disk space, device name changes, and system load). After participants reviewed and signed the consent form, all cameras were activated. Participants were asked to wear pinhole camera eyeglasses; those with their own glasses either wore both or removed their personal pair if comfortable. A movie clapboard was used to synchronize all video streams. When the participant felt ready, they began the first task. After each task, the cameras were stopped (partly to reduce the load on the recording system) and the experimenter switched to the next machine, then restarted the cameras and repeated the clapboard synchronization. During the tasks, the experimenter remained behind the setup, only intervening for the pen plotter task (by controlling AxiDraw from a laptop) and when a participant struggled with using the sewing machine (for safety reasons). At the end of the three tasks, the participant took part in a brief, audio-recorded interview while still seated. Once the session ended, the experimenter photographed or scanned each completed artifact (clay print, sewn fabric, watercolor painting) and disinfected the eyeglasses before the next participant. We audio-recorded and automatically transcribed interviews using Whisper [OpenAI, 2022] on a local GPU, with manual corrections.

4.4.4 Participants

Twelve study participants were recruited from Cornell Tech using e-mail circulation, direct approach, and snowball sampling. Participants generally had little or no experience with each machine. Experience was lowest with the 3D clay extruder as only P9 and P10 reported having some experience with similar extruders. Three participants had had an initial experience with a pen plotter (P5, P10, P11) while one participant said that they had built a drawing machine in the past (P12). Three participants said that they had tried sewing machines in the past (P3, P10, P12) with one participant mentioning having two years of on/off experience on the sewing machine (P7).

4.4.5 Data Processing

We synchronized the data streams using the audio-visual signal from the movie clapboard, and trimmed the footage length to match the task duration. We then created a combined view incorporating both the eyeglass camera footage and the 360° footage (either split into front and back views or presented as a single wide-angle view for the second task). This resulted in 3 hours and 48 minutes of data per camera, across the total of study participants and tasks.

4.5 Analysis

We analyzed the study data through two complementary approaches: automated body pose detection to quantitatively assess camera observability, and exploratory qualitative coding to identify the types of interactions and intervention opportunities captured across camera views.

4.5.1 Quantitative Analysis

To assess camera observability across participants and tasks, we conducted automated body pose detection using OpenPose [Cao et al., 2018] with the BODY25 model and face tracking. We processed all video frames from the task performance periods (3 hours and 48 minutes total across all participants) for the participant-facing cameras: the ZED 2 front camera, the 360° participant-oriented view, and the left and right overhead webcams.

For each frame, we recorded which of the 25 body keypoints were successfully detected. We excluded lower-body keypoints (legs and feet) from analysis as these were consistently occluded by the table across all camera angles. We computed detection rates as the proportion of frames in which each keypoint was identified, filtering to include only frames where the participant was the sole detected person to avoid confounding effects from experimenter presence. We aggregated detection rates across all study participants and tasks to create observability profiles for each camera angle. This allowed quantitative comparison of which camera perspectives most reliably captured participant body pose, providing an objective measure to complement qualitative observations about interaction content. The results of this analysis are presented in Figure 4.7.

4.5.2 Qualitative Analysis

As an exploratory first pass on the research data, one researcher, who was also the experimenter in the room, analyzed the combined footage by marking moments of interest (i.e., perceived moments where there was either a task failure or another opportunity for a machine to interact and offer assistance) in ATLAS.ti. We dynamically updated the labels according to the emerging categories of situations and signals. The coding process led to the identification of 23 preliminary code categories, with a total of 359 labeled instances.

The interactions observed while participants performed the three tasks on the three machines were rich and diverse, giving a broad picture of potential opportunities for interaction that could be observed from the captured audiovisual data. The code labels and their frequencies are presented in Table E.1. In parallel, we discussed selected moments of interest repeatedly among members of the research team. The findings from this exploratory analysis are discussed together with our analysis of the instrumentation setup in order to evaluate the appropriateness of multi-view camera instrumentation for interaction analysis. The findings are grounded in patterns that arose in the interview data.

4.6 Results

In this section we revisit how our study informed our two research questions: whether multi-view camera instrumentation can serve as a practical method for capturing human-machine fabrication interactions, and what these observations reveal about the signals and situations that characterize these interactions. We address these questions for each individual machine, but begin with a higher-level discussion.

4.6.1 Interaction Vignette

Throughout our study we found that the situations that unfold can quickly become highly complex. To illustrate this, we visualized an example excerpt in Figure 4.6 showing just the 360° and egocentric camera views. Without lowering the presser foot, a novice study participant attempts to sew from right to left. The sewing machine gets stuck. The participant looks at the side of the machine to see what is causing the problem and tries to lower the presser foot, but sees that the bottom part of the foot has come off as he turned the wrong lever earlier. The experimenter tries to come in to explain what happened, not realizing that the foot has come off: "So when... oh, this one came off." The participant asks: "What did I do?" The experimenter tries to re-attach the foot. Participant: "Oh my god." Both laugh. Participant: "I think I shouldn't use this machine." Key observable indicators of breakdowns in this sequence are machine state signals (the foot falls off and the presser foot is not lowered) which are seen visually through the egocentric camera. We also see that the participant starts to look

Figure 4.6: An example of the complexity of interactions: here, a novice tries to sew, but mishandles the machine, leading to a part falling off. He laughs off the mistake and gives up. A combined egocentric and machine-mounted 360° view with audio gives the video analyst sufficient information, while preventing a visual overload. However, room cameras could add additional visual cues regarding machine handles that were misconfigured.

closely at the point of the presser foot, apparent due to posture changes that can be seen in the 360° machine-mounted camera, as well as through a tilt in the egocentric camera view. Increased apparent size of objects of interest in the egocentric view indicates a participant looking closely. We also hear speech and utterances indicative of nervousness, as well as machine sounds indicating the foot pedal being pressed, but we do not see the foot pedal. This excerpt shows the complexity of the observable signals that indicate a breakdown.

No single camera perspective captured all signals at once. The egocentric view showed the participant's gaze and their actions, but not their facial expression. The 360° camera captured body pose and facial expression but offered a limited view of the working area due to occlusion by the machine. Audio revealed machine sounds and emotional tone, and room cameras, although not visualized, could add additional viewpoints to see the mechanical failure at the presser foot mechanism.

This example illustrates that effective interaction analysis for fabrication machines requires capturing: (1) machine actions and state, (2) material and/or artifact state, (3) human actions and affect, and (4) contextual information of the larger work area.

4.6.2 Camera Observability

Every task was instrumented with 5 cameras, leading to 5 or 6 video streams depending on how the 360° data was exported—for the pen plotter, we exported a combined wide angle perspective rather than front-and-back due to the location of the camera on the machine (see Figure 4.3). Across these camera angles, the machine at times occludes the view of the person's body.

To analyze to which extent the observability of participants generalizes across participants and tasks, we took the entire dataset and used the availability of body pose keypoints as a proxy for observability (i.e., the view from that particular camera is neither occluded nor too limited to observe the participant). The results are displayed in Figure 4.7. The results were obtained by processing the data during which tasks were performed (i.e., 3 hours and 48 minutes, see Section 4.4) through OpenPose [Cao et al., 2018]. We excluded the eyeglass camera data and the A2 (360° angle oriented away from participant), as these cameras were not oriented towards the participant and thus not relevant for pose visibility.

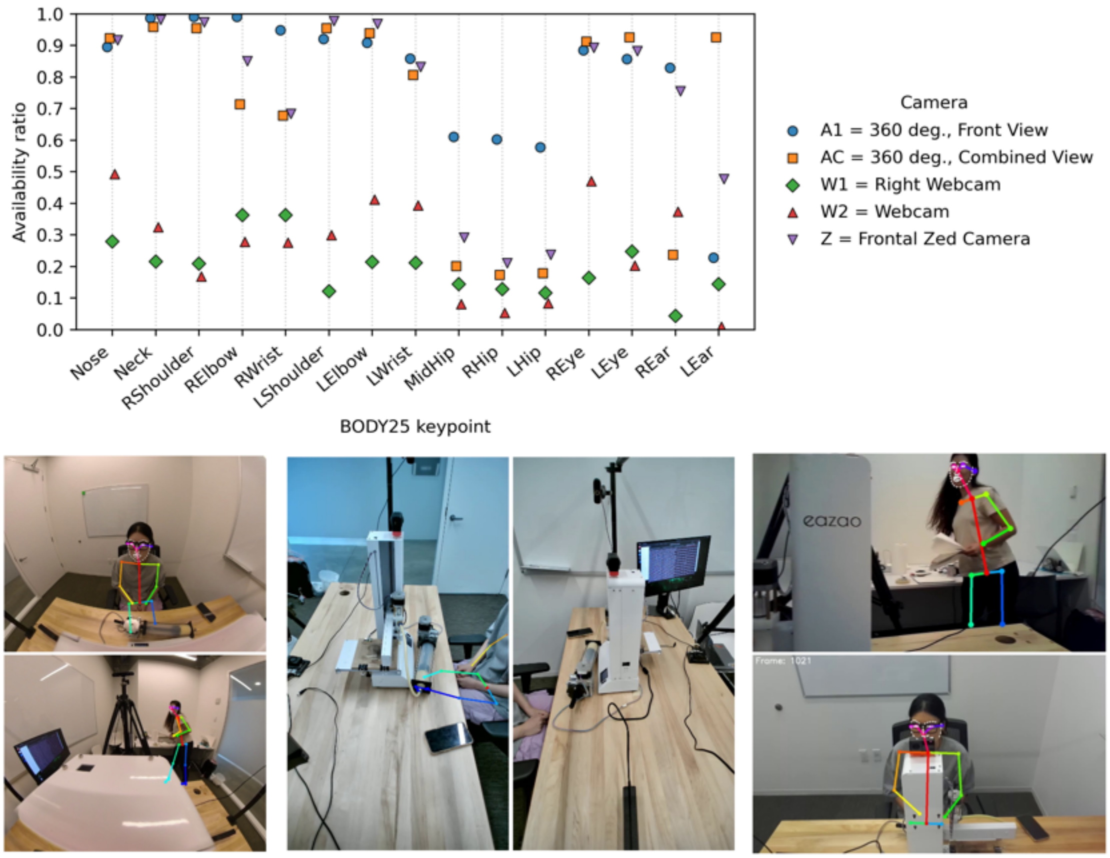


Figure 4.7: (Top) Availability of each BODY25 keypoint (legs excluded) over all frames where only the study participant is in view ($N_{\text{persons}} = 1$, aimed at participant). (Bottom) Overlay images showing different camera views, here for the clay extruder. Note that, in the pen plotter condition, A1 and A2 are always replaced by the combined AC view.

We show the availability ratio for each keypoint (excluding legs as they are occluded by the table) at the top of Figure 4.7, and visualize the BODY25 keypoint overlay for each of the camera feeds at a specific moment in time at the bottom of Figure 4.7. From this analysis, we can see that the 360° participant-oriented machine mounted camera performs consistently well for almost all keypoints—the performance dips for below-waist keypoints which is expected due to the location of the table. As we see in the overlay images, the left and right webcams were focused on the machine rather than on capturing the entire body of the user. Whereas in overall counts they did not lead to better observability than the machine-mounted camera, there are specific moments in the dataset (such as the moment in Figure 4.6), where having a view from the side of the machine, if needed, could provide extra contextual information into machine states. During the coding process, however, we found that observing three video feeds at once was a good tradeoff between allowing the observer to see what is happening, while not overloading them with information.

In the following section, we will cover the findings for our two research questions separately for each of the three machines, starting with the clay extruder, then the pen plotter, and finally the sewing machine. For each machine, we list out the interactive affordances, the types of failures that could occur, and then evaluate how this was visible through each of the cameras in the setup. In addition, we briefly discuss participants' overall subjective experiences.

4.6.2.1 Clay Extruder: Machine-Controlled Motion, Machine-Controlled Material

The clay extruder task represented a highly automated fabrication process with an unpredictable material. Study participants attempted to create a cylinder from wheat-and-corn dough, a task deliberately designed to have a high probability of material failure. This design choice allowed us to observe how participants monitored, interpreted, and responded to machine behavior when direct control was minimal.

The Eazao Zero clay extruder (depicted in Figure 4.4) featured a touch screen on the right, for machine control and the loading and starting of print files. Physical clips kept the MDF printing bed in place, and the extruder nozzle moved in three dimensions relative to the printing bed while depositing material. Once a toolpath was uploaded and initiated, the machine operated largely autonomously unless it was stopped or its settings were changed through the touch screen.[5]

Interactions occurring around the clay extruder during the printing phase could be supervisory (e.g., monitoring progress and interpreting visual and auditory cues) and interventionary (e.g., pausing the machine or touching the clay in the case of an error). Unlike the other two machines, the clay extruder offered limited real-time control as the machine was both in control of the material and in control of the movement path. Study participants could interact with the machine through the touch pad (e.g., pausing the machine), or manually manipulate the printbed or the dough (e.g., correcting small defects), but they could not easily adjust the path or extrusion settings as these either required a computer or required multi-layered navigation of the machine interface which they were largely unfamiliar with. The additive nature of the task meant that errors compounded over time rather than being easily correctable.

Common issues that are expected with clay printers include problems with the machine (wrong settings, unleveled printing bed, wrong choice of nozzle),

[5] The machine was also operable via a computer over a WiFi connection, but that functionality was not used during the experiment.

with the material (bubbles in the clay, too wet or too dry clay, inconsistent mixing, not enough clay), and with the print (unintended structural collapse). In our studies the most common issues related to the material's variable consistency and suboptimal print settings that resulted in a failed print. Dough that was too dry resulted in slowed extrusion and poor layer adhesion; dough that was too wet frequently led to structural collapse. The first layer also frequently failed to adhere to the print bed, causing subsequent layers to collapse. Even when the first layer was successful, the layer height setting was slightly too low for the print to succeed, and the nozzle would travel through the last layers of the print, often causing either collapse or deformation. The slow, layer-by-layer build process meant that problems developed gradually, creating anticipatory tension as study participants watched for signs of impending failure. Participants varied in their responses to these failures: some participants attempted to correct adhesion issues with the provided tool and used their hands and/or a tool to reshape the cylinder after prior issues, whereas other participants did not intervene.

The egocentric camera succeeded in capturing what the study participants were attending to in most cases; however, for some participants the field of view of the camera was located slightly higher than the area that they were looking at. The egocentric camera also captured hand movements when participants intervened, such as using the tool to adjust the clay in place. However, the egocentric view had limitations: it did not capture the participant's facial expressions or broader body posture, and when participants looked away from the machine (e.g., toward the experimenter), this camera lost visibility of the machine and artifact state.

The 360° machine-mounted camera gave a good view of study participants' facial expressions and body pose. This camera also captured actions that participants took and their manual interactions with tools, with the machine, and with the material.

As the machine had a "pillar" on the right side, the left side webcam could sometimes provide a view of the artifact that was not available from the 360° camera (due to machine occlusion) nor from the eyeglass camera (due to participant's gaze direction). The ZED 2 front-facing camera duplicated much of what the 360° participant-facing view captured, but from a more distant vantage point and featuring occlusion by the machine.

Thus, for the clay extruder, the combination of egocentric and machine-mounted 360° (participant-facing view) video mostly captured the essential signals: artifact state and participant attention (egocentric), plus facial affect and body pose (360° camera). Room cameras added marginal value for this task, with the left-side camera at times adding visibility to the artifact. No camera captured the internal machine state (e.g., extrusion speed settings), but these data could be obtained through additional machine logging. In addition, the cameras could not directly measure material properties that influenced print success, such as moisture content and viscosity.

4.6.2.2 Pen Plotter: Machine-Controlled Motion, Human-Controlled Material

The pen plotter task featured highly predictable machine behavior combined with unpredictable (due to watercolor behavior), but human-controlled material behavior. Participants controlled watercolor paint and water application while the AxiDraw V3 precisely executed a pre-programmed spiral path. This created a human-machine collaboration where success depended on both parties' contributions.

The AxiDraw V3 is a 2.5D machine that can move a writing instrument on a piece of paper and lift it either up or down. The pen plotter typically follows a precomputed path, and regular usage of pen plotters has a high rate of success due to limited variability in the material. Users typically prepare a drawing file (either visually with a software extension such as Inkscape [Evil Mad Scientist Wiki, 2024], or programmatically [Evil Mad Scientist, 2025]) and can observe the drawing unfolding in real time. Despite high precision, unexpected results can occur due to misalignment, pen faults (e.g., ink running out), or mechanical issues causing drift. In our study, the plotter held a watercolor brush (kept empty of built-in water) and traced an outward spiral path. Since the print was started by the experimenter, the affordances of the plotter were limited to the following: the sponge could be squeezed or taken off to control water flow, paint and water could be added to or removed from the paper (using hands, the sponge, or paper towels), and the path could be modified by moving the plotter altogether (an unintended affordance that we observed in the dataset).

Unlike the more goal-oriented clay extruder task, issues with this machine depended largely on participants' individual preferences and judgments. Common issues included: adding not enough or too much water, unintended marks on the paper, a mismatch between expectations of where the plotter would go and where it would actually go, and spatial constraints due to the clips holding the paper in place (as seen in Figure 4.9). Importantly, most participants reported that they did best on this task, with some reporting that the plotter

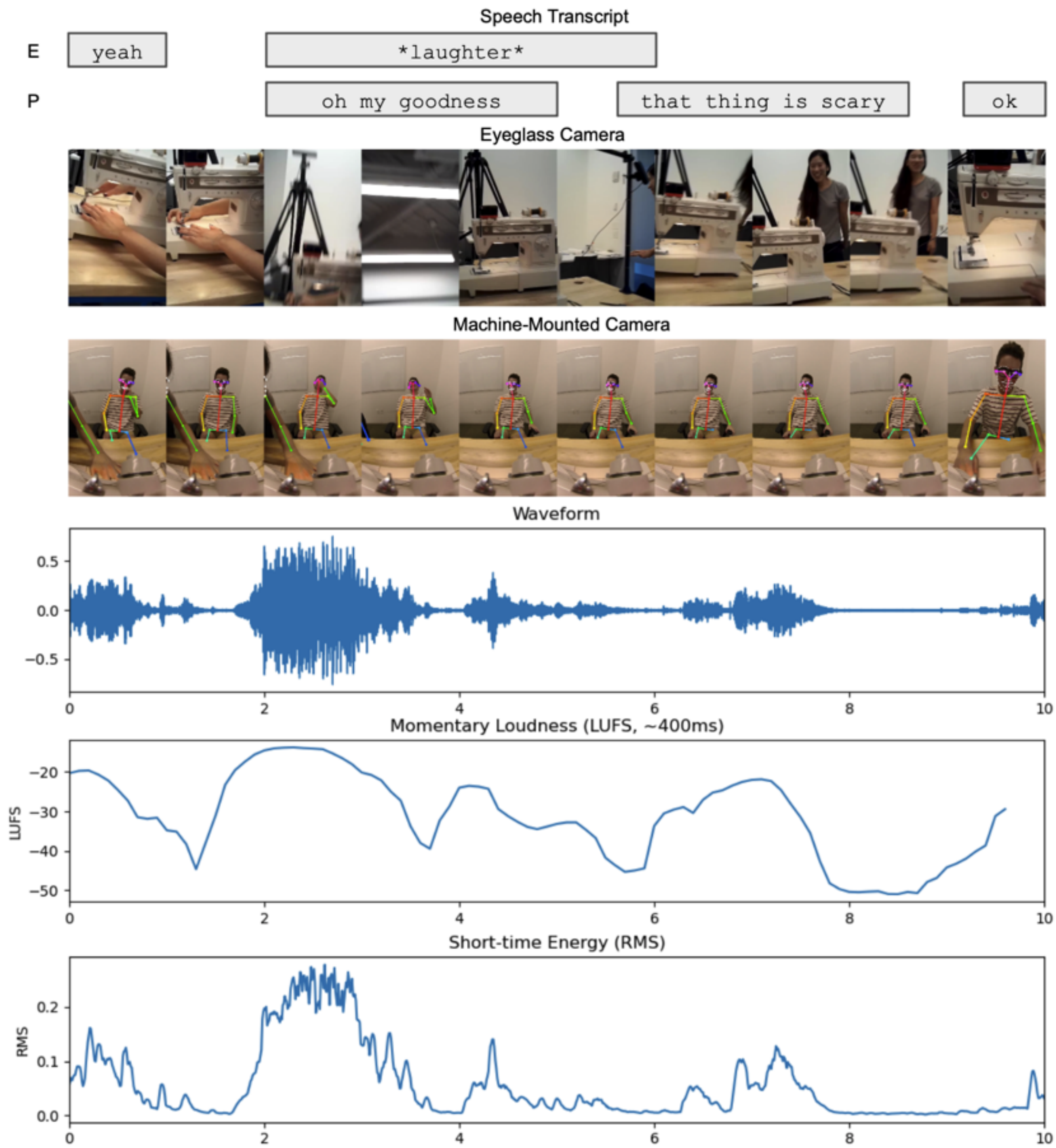


Figure 4.8: A 10-second sequence after P11 asks the experimenter for help. The experimenter ("E") presses the foot pedal, startling P11 ("P"). This is apparent from the eyeglass camera (looking away at the ceiling), body pose data (moving away from the machine), and the audio track (machine action causes loudness).

helped their creativity (see Section 4.6.3).

The egocentric camera perspective was helpful for capturing hand-material interactions, such as adding paint or water or using paper towels. This camera also gave insight into study participants' attention patterns as the field of view followed participants' gaze. Due to the plotter's low profile and the camera's mounting position, we exported a single wide-angle view rather than separate front/back perspectives for the 360° camera. This view showed participants' (upper) body pose, facial expressions and captured all their movements in the workspace. With the profile of the plotter, the room cameras added limited value, aside from being able to capture the work area in cases where the participant was looking away and the egocentric camera perspective would not suffice. Audio signals were useful to capture participants' speech and utterances, but, in contrast to the clay extruder and sewing machine, the sound of the servo motors on the pen plotters was relatively stable and less indicative of errors. Using cameras it was not possible to capture the exact wetness of the paper or the exact amount of paint that was present on the paper.

4.6.2.3 Sewing Machine: Human-Controlled Motion, Human-Controlled Material

Sewing machines are 2.5D machines that can move thread on and through a piece of fabric, for example to attach two pieces of fabric together. Our sewing machine presented the most complex interaction dynamics: continuous human control via foot pedal, numerous mechanical settings and failure modes, and a fully analog interface.

The sewing machine had no initiative, as most of the sewing machine's actions are directly controlled by the human user. The user operated a foot pedal to control speed, guided fabric through the feed dogs, and maintained fabric alignment by hand. Manual control for precise single-stitch advancement was available through the handwheel. During a typical sewing process users would constantly adjust posture, grip and pressure. Pauses were used to realign or reorient fabric, inspect stitches or solve thread jams. The sewing machine had a number of dials and settings that could be adjusted, but that did not need adjustment during the study. Furthermore, it had a wheel on the side to manually control the stitches. The participant also needed to interact with the lever to raise the presser foot, and was instructed to use scissors to cut thread once a seam had been made.

During the experiment, the sound of the machine would be a first factor to give away whether or not any stitches were being made. The loudness of the machine indicated the force with which the foot pedal was pressed, and indirectly the confidence of the participant. Insecure participants tended to touch the foot pedal very carefully and get startled when the machine started moving (see P11 in Figure 4.8). Furthermore, changes in the participants' body pose, such as moving in order to get a better view of a different part of the machine, also informed whether the likelihood was high that they needed help with the task (see P1 in Figure 4.6, who, especially in the second frame, changes his body pose to get a better look at the machine).

The sewing machine presented the most demanding observability requirements: simultaneous monitoring of hand positions, fabric state, machine configuration, foot pedal control, and affective state. We illustrate a sequence and the audiovisual information streams to illustrate the richness of each signal. In Figure 4.8, P11 has just asked the experimenter for help. He made some stitches using the sewing machine, but arrives at a point where pressing the pedal all the way down only results in a motor noise, but not in the machine continuing to make stitches. The experimenter comes in and presses the foot pedal, resulting in a louder noise that startles the participant and causes them to say 'oh my goodness, that thing is scary' while both laugh. We see the ten seconds from 4:04 to 4:34 of the synchronized video streams visualized at the top, taking one frame at the start of each second (cropped for visibility). From second 2 onward, a loud noise is visible in the audio stream. Looking at the eyeglass camera at this point, we see the participant looking up to the ceiling before looking back at the machine and the experimenter. This movement is confirmed in their body pose changes as seen from the machine-mounted camera.

In general, the egocentric camera perspective was able to capture participants' visual attention and hand movements, although this wasn't the case when participants used the manual handwheel without looking there. The 360° perspective was able to capture body pose and facial expression, but the view of the working area was largely occluded due to the position of the camera being on top of the machine. Room cameras added valuable perspectives due to controls on the machine being out of the view of the user, such as the lever at the back of the needle and the handwheel on the right side of the machine (see Figure 4.3 for viewing angles from each of these cameras). The ZED camera was inferior to the front view from the 360 camera, due to occlusion by the machine. None of the cameras captured the foot pedal, but sounds of the machine gave an indication of whether the pedal was pressed and with how much force. The cameras could not capture internal machine states (thread tension, bobbin thread remaining), or the precise force applied to controls. Machine logging and force sensors could complement visual data.

4.6.3 After Task Comments

Post-experiment semi-structured interviews revealed study participants' subjective experiences during the tasks as well as their interpretations of their interactions. We present the findings grouped by question topic. The list of questions asked is given in Appendix D.

When after the experiment we asked participants, "What are your initial thoughts?", most participants remarked that they enjoyed the tasks, mostly because the machines were novel to them (P6, P7) or because they liked creating things (P9, P10). For P8, the novelty of the machines motivated them to succeed in the tasks. P7 remarked that the experiment setting encouraged using a difficult machine: "Sometimes I give up easily when it comes to sewing. So it was nice to have a structured environment where if something goes wrong, I don't feel lost and alone." P1, P2, P3, P4, expressed specifically enjoying the pen plotter task, with P3 stating: "I really liked [it] because it was easy to intervene in the creation process." P10 found that the pen plotter helped them: "As you get older, it becomes more difficult to come up with ideas. But since the machine kept moving, it was a spice for my creativity." P3 remarked on feeling limited ability to control or stop the clay extruder or pen plotter: "For the clay extruder, it would go around no matter what. Even where I felt it definitely should stop. It was messing up and I'd try to fix it, and it would only make it worse." Similarly, P12: "Because it's programmed to go layer by layer, once it breaks, there's no way to save it." Most participants found the clay printing task the most difficult, because the machine's intent and behavior were hard to understand or control (P3, P4, P7). P11 remarked on the difference between machines: "[the sewing machine] is like a tool, whereas with the other two it's more like a collaboration." Participants were also asked if they experienced specific moments of joy, difficulty, confusion, or frustration.

The participants varied in their self-rated task performance. Three participants described their own performance as bad on all three machines (P1, P3, P11). Participants were most likely to view their performance on the drawing task as neutral or good (P3, P4, P5, P6, P7, P8, P9), which was mostly explained as stemming from the open-endedness of the task and the predictability of the machine movement. The extruder task was perceived as offering no opportunity for intervention or flexibility (P4, P7). Participants also perceived the extrusion task, which frequently resulted in a collapsed cylinder, as a failure of the machine (P6) (Figure 4.9) or inadequacy of the material (P10). However, one participant stated that they did best at the extrusion task because they fixed the machine's failure (P11). Regarding the sewing task, P6 said their lack of experience influenced the outcome, whereas P7 was happy with their performance.

Figure 4.9: A selection of creations from the study. Our study participants' approaches to tasks differed, as did their self-rated task performance.

When asked what, if anything, they would like to change about the study setup, tools, and machines, P1 and P3 mentioned that they would like to be more involved in the setup and preparation of the tasks. As stated by P1: "with the drawing, I would want to see what shape and where it should start, because I was confused. I put the color on the paper beforehand and it started at the side, so I should have pinpointed, maybe marked the paper or do something to know that it was going to start circling from this point." P3 and P10 suggested a change in orientation of the extruder, such that they would have a better reach and would not have to avoid colliding with the extruder while touching the clay. P6 mentioned that they liked interacting with the machines at eye level, although if they would be using the setup for a long period of time they might need a posture change. Some participants mentioned that they might have liked to stand in the painting (P7, P12) and extruding (P7) activity, and that the clips on the paint paper were obstructing their reach. Furthermore, P7 also mentioned that they would like to use the machines simultaneously as the processes were quite slow.

4.7 FabriCam-5 Dataset: Preparation for Future Analysis

As a contribution to the research community, we prepared the FabriCam-5 dataset for public release on the Harvard Dataverse [Bremers et al., 2025]. The dataset consists of synchronized multi-view video recordings from our study of 12 participants interacting with three fabrication machines.

4.7.1 Dataset Contents

The dataset includes five synchronized video streams per participant-task combination: two side webcam perspectives (left and right), one front-facing ZED 2 camera view, one egocentric pinhole camera perspective, and one 360° machine-mounted camera view (extracted for front- and back perspective for the sewing machine and clay extruder, and a single side perspective for the pen plotter). Furthermore, it contains audio captured from the eyeglasses camera and the 360° camera.

4.7.2 Data Processing

All video streams were temporally aligned using the clapboard synchronization signal recorded at the start of each task. Video was trimmed to include only task performance periods, excluding setup time and interview segments. The 360° footage was exported either as separate front and back views or as a combined wide-angle view, depending on which format provided better visibility for each machine (see Figure 4.3). Participant faces are visible in the dataset, except for blurring where requested. Participants provided informed consent for data sharing as part of the study protocol, which was approved by the Cornell University Institutional Review Board.

4.7.3 Intended Uses

The dataset is intended to support research in human-machine interaction, video-based interaction analysis, and machine learning for behavior recognition in fabrication contexts. Potential applications include development and validation of automated detection methods for the interaction situations and signals identified in this chapter; training of models for recognizing affective states, errors, or opportune moments for machine assistance; comparison of instrumentation approaches against our documented multi-camera setup; and analysis of novice behavior patterns across different fabrication modalities.

4.7.4 Dataset Availability

The dataset will be made available upon publication of the associated paper, at the Harvard Dataverse. Researchers using the dataset should cite both the dataset record and this dissertation or its follow-up publication.

4.8 Discussion

The instrumentation study demonstrated that a lightweight combination of egocentric and machine-mounted cameras captures the majority of relevant interaction signals during fabrication, and that the interactions themselves are richer and more multi-modal than a focus on task performance alone would suggest. This section discusses what these findings mean for our research questions, proposes a generalizable instrumentation protocol, and considers implications for the design of initiative-taking fabrication machines.

4.8.1 Evaluation of Research Questions

Through a study with 12 participants using three fabrication machines, we investigated whether multi-view camera instrumentation can serve as a practical method for capturing the interactions that inform the design of human-machine collaboration, and what these observations reveal about the nature of fabrication interactions.

We found that human-machine interactions involved rich, multi-modal temporal signals including human behavioral and affect states, machine states, and artifact states. We further found that a combination of egocentric and 360° machine-mounted cameras captures the majority of relevant interaction signals related to the artifact, machine and human, while still being practical and light-weight to collect and analyze. Adding specific room cameras could improve the visibility of machine and artifact to compensate for machine-related occlusions and tasks performed outside of the user's gaze direction.

We address our first research question, on camera instrumentation, before turning to our second, on what these observations reveal about fabrication interactions. In addition, we propose a protocol for instrumenting fabrication machines including practical guidance on camera selection based on research goals, and explore implications for designing collaborative machines.

Our first research question asked whether multi-view camera instrumentation could effectively capture the signals we identified as essential. We found that a combination of two perspectives, being the human's perspective and the machine's perspective, is able to capture most important interactions. While the intuitive approach might be to set up a camera at an angle outside of the working area, there isn't a natural place for the camera to go without taking up the perspective of a third person observer, leading to occlusions and distance from the unfolding interaction. The combined human-centered and machine-centered views provided a comprehensive coverage of the embodied activity and expressions during the interactions.

The egocentric perspective captured what the human saw, touched, and attempted; both when it came to the machine or artifact, and when it came to the contextual space where the interaction unfolded. The machine-centered view captured the user's facial expressions and body pose, as well as a surround view of the controls on the machine and its surroundings. Unlike external room-level views, which often suffer from occlusion, a lack of detail, or a bias in the choice of their placement, the egocentric and machine-centric 360° placement allowed for a systematic, logical placement of cameras based on the two parties to the interaction. Aside from that, placement of a single 360° camera on a machine also comes with practical benefits, as external camera placement would also lead to additional requirements on the space where the machine is being used. The setup could benefit from additional room cameras when machine controls are located in poorly visible locations, and when analyzing users whose gaze might not follow hand movements (such as experts operating the sewing machine).

There are practical considerations that should be kept in mind when selecting camera setups. First, during video analysis, we found that viewing three synchronized streams was manageable and provided comprehensive coverage without becoming too overwhelming to the analyst. For coding and analysis we recommend limiting simultaneous viewing to 3 streams at once, and using additional camera angles as supplementary references as needed. Second, 360° camera footage typically results in large files that require more complex post-processing than other types of cameras. These aspects should be kept in mind while piloting an instrumentation setup. Having established that multi-camera instrumentation provides sufficient coverage, we now turn to our second research question: what do these observations reveal about the signals and situations that characterize fabrication interactions?

To organize our observations, we draw on a framework of four interconnected signal categories. These categories are not novel in themselves; they follow naturally from the structure of any human-machine interaction involving physical materials. Their value here is in specifying what each category looks like in the context of fabrication, and in identifying which signals within each category are most relevant for informing machine initiative. The action space consists of four interconnected categories in which signals can fall: human state, machine state, artifact state and contextual state. These signals can often be observed as discrete events, but they are best understood by analyzing their evolution over time during a task.

Human state signals include both implicit and explicit interaction signals that indicate cognitive and affective states and behaviors. Explicit signals include, for instance, verbal utterances (e.g., exclamations, self-talk), and goal-oriented hand movements (e.g., adjusting machine controls, picking up tools). Implicit signals can include facial expressions (e.g., a frown revealing frustration or concentration), body posture (e.g., moving closer to get a better look, or moving away due to being startled), and gaze patterns (e.g., indicating visual searching). We observed that implicit signals often preceded explicit help-seeking behaviors, which suggests opportunities for proactive assistance by an observing machine. This pattern was most evident during sewing machine interactions, where the task required continuous human operation and thus created more opportunities for difficulty to arise. In nearly all cases where participants struggled with the sewing machine, visible hesitation, postural shifts, or changes in gaze direction preceded verbal expressions of confusion or explicit requests for help. With the pen plotter and clay extruder, where participants often adopted an observational role, these implicit behavioral cues were less frequent simply because participants were less actively engaged with error recovery around the machine. Task state itself also carried information about the need for intervention: a collapsing clay print or a misaligned stitch provided signals that fall somewhere between implicit and explicit, readable by an observer but not deliberately communicated by the participant. Together, these observations motivated the design of the Wizard-of-Oz study in Chapter 5, which used the sewing machine because it demands sustained human engagement, and in which the wizard monitored participants for implicit behavioral cues and task state changes to deliver proactive assistance.

Machine state signals of interest include operational status (e.g., is it running, paused, or broken down), sounds indicating performance (e.g., normal motor sound or strained sound), and configuration settings (e.g., extrusion rate or plotting speed). While some machine states are directly observable through camera instrumentation (e.g., mechanical positions of dials and levers), others require supplementary data streams such as machine logs, sensor readings, or API access. For our selected clay extruder and pen plotter such logging would be feasible; for our analog sewing machine, computer vision or embedded sensing approaches could be leveraged to infer machine state. In addition, machine sounds often provided a rich source of information, in particular for the sewing machine where loudness correlated with foot pedal pressure (see Figure 4.8).

Artifact state signals relate to the product that is being created during the human-machine collaboration. For the clay extruder, important signals would include layer adhesion quality, structural stability, deformation, and collapse. For the pen plotter, artifact state included paint/water distribution, color levels, and eventually, emergent shapes and patterns on the visual paper. For the sewing machine, artifact state includes stitch formation and fabric alignment. Artifact state was often a trigger for participant interactions either due to errors that they tried to recover from (e.g., print collapse), or as a prompt for interaction (e.g., shapes on the paper inspiring the addition of paint in a certain area). This aligns with Schön [2017]'s statement that in reflective practices, a person is in conversation with a material and responds interactively to the material.

Contextual state signals relate to the wider environment in which the interaction is taking place. Aspects here can relate to workspace layout, availability of tools, and the presence of other people. Due to the controlled nature of our lab study, the contextual state remained relatively stable, but we envision that contextual state will play a larger role when similar data is collected in (often shared) natural making spaces.

In our study we observed several types of breakdowns: machine configuration errors, material failures, human-machine coordination challenges, and human affective breakdowns. In line with foundational works in the CSCW and HCI literature, breakdowns are often valuable moments of insight (see Suchman [1987]). Often, the richest moments in our study involved unexpected, subtle human behaviors and contextual signals occurring during breakdowns or moments of uncertainty. Participants frequently showed expressive movements, like leaning in closer, tilting their head, or pausing or retracting a hand, when they encountered an unexpected situation. These nuanced observations could be early indicators of problems or suboptimal human-machine interactions. Exactly these unexpected (and, at times, undesirable) moments where user intent and machine behavior are misaligned, offer opportunities for interaction design to be introduced to mitigate breakdowns.

Examples of machine configuration errors included forgetting to lower the presser foot on the sewing machine. We limited the configuration required by users of the machines in our study, since the clay extruder and pen plotter were both executing pre-programmed paths and the sewing machine was already configured prior to the interaction. Material failures included unexpected material behavior despite correct machine configurations, such as the print collapse in the clay extruder task caused by material that was too moist. Coordination challenges arose when the difficulty was mainly in the synchronization between human actions and machine actions, such as participants trying to time the addition of paint without colliding with the machine in the pen plotting task, and difficulty finding the right pressure to use on the foot pedal of the sewing machine. At times we also observed affective breakdowns, such as observed nervousness, frustration, and loss of confidence. These states were often caused by repeated failures or by encounters with unpredictable machine behavior.

Due to the open-ended nature of creative tasks, task failures cannot always be objectively determined by looking at task outcomes alone. As we found in our interviews, the attribution of failures varied (e.g., some participants attributed a collapsed cylinder to a failure of the material, whereas other participants stated that they did well because they successfully repaired a collapsed cylinder, see Figure 4.9). The source of enjoyment is also variable: whereas some participants actively engaged with the materials, others explicitly noted enjoying passive observation of the machine. This suggests that machines that support creative work should avoid imposing narrow success criteria, and instead try to adhere to the users' goals and experiences. Similarly, the process of attempting to learn machine operation could be valuable even if the final artifact has flaws.

Each breakdown type offers specific opportunities for machine intervention towards repair or support. As we saw in our study, systems for capturing interactions should not only support the planned interaction, but also account for and offer enough flexibility to be able to capture emergent and unpredictable realities of creative practice. We found that multi-camera instrumentation was able to capture a rich variety of failures. To ensure the highest likelihood of these unexpected situations being captured, interaction designers can refer to our proposed instrumentation protocol as outlined in Section 4.8.2.

4.8.2 Proposed Instrumentation Protocol

Based on our experience and findings, we propose a systematic protocol that interaction researchers can follow to determine appropriate instrumentation for capturing (tabletop) human-machine fabrication interactions. This protocol is designed to be generalizable across different fabrication machines while allowing customization to specific research contexts.

First, it is important to outline the affordances and characteristics of each machine, in order to determine which essential aspects should be captured. This could include control interfaces and blind spots created by the machine itself.

Second, research questions inform signal priorities. For instance, interaction elicitation studies might emphasize the capturing of human aspects of the behavior, over capturing the quality of the resulting artifact. Data collection studies for training machine learning models to perform tasks might, in contrast, require a higher level of fidelity of artifact and machine state data. In the case of a Wizard-of-Oz study, the ability to view camera data in real-time matters. In these cases it could be beneficial to select cameras with robust streaming solutions to prevent issues related to signal disconnection or overheating.

Third, instrumentation should be guided by an informed estimation of where moments of transition, hesitation, or recovery are likely to occur, as these are typically moments where rich design insights can be derived. While these moments cannot be predicted with certainty, machines can come with documented frequently occurring errors in the user manual. We recommend that experimenters select a few of these errors, to be included in their setup testing protocol, even if an experimenter is only pretending to make a certain error.

Fourth, we recommend starting with the minimal combination of the egocentric perspective and a 360° machine perspective. Hardware should be selected with the nature of the task in mind, balancing data quality with ease of operation and mounting and level of obtrusiveness to the user. Specifically regarding the user-centric camera perspective, glasses can be considered, but they can conflict with the user's own eyeglasses if they need to wear them. Furthermore, differences in head movements and facial structure can cause differences in the final field of view, and it is recommended that experimenters test this with a range of different people before narrowing down on a final specific sensor to record the human-centered perspective. For the machine-mounted camera, we found that velcro tape provided a relatively strong yet impermanent mounting solution that allowed precise control over the mounting location.

Fifth, we recommend piloting the setup by recording various people performing the tasks naturally starting with two camera angles, and then to deliberately enact various types of edge cases and mistakes that could occur, to be able to verify that the recording setup is able to capture them. Specific attention should be paid to the visibility of the relevant machine controls and artifact states, as these can easily become obstructed from view by the decision of where the 360° camera is mounted. From recordings, algorithmic analysis of the resulting data streams could reveal occlusions (e.g., controls near the side or back of the machine that cannot be easily seen from the 360° camera). We recommend that edge cases are deliberately included during pilot testing, such as configuration errors, such that breakdown moment capture can also be evaluated. Our scripts to arrive at body pose occlusion evaluation depicted in Figure 4.7 will be made available on a GitHub repository.

Finally, blind spot analysis can inform the addition of camera angles to solve for the specific occlusions, and iteratively the ideal setup can be arrived at.

4.8.3 Design Implications

Our findings have implications for designing future interactions where signals could be leveraged by machines to improve collaboration in creative tasks through multi-modal sensing and mixed-initiative collaboration. In addressing which signals future machines might leverage to become better collaborators, our study discussed signal types that appear to be relevant from our preliminary observations: human state (e.g., body posture, facial expression, gaze), machine state (e.g., operational status, feedback mechanisms), artifact state (e.g., visible progress or failure), and contextual state (e.g., presence of other actors, tool locations, space where the interaction is taking place). Video instrumentation from the perspectives of the human (eyeglass) and the machine (360°) provides for data that can be algorithmically processed to derive data streams such as gaze, pose, utterances, as well as scene understanding and object detection when it comes to tools, machine and artifact. These signals develop over time and have a temporal nature: rather than reacting to discrete signals and events, evolving combinations of signals could be interpreted in relation to task progress, user history, and context of the interaction.

The signals we identified as observable through camera instrumentation provide opportunities for machine intervention. For example, machine configuration failures could trigger proactive reminders by the machine to check settings and material failures could trigger practical guidance and encouragement. Coordination difficulties could, depending on the machine, be reduced through adaptive settings, such as a slower plotting speed or by offering tutorials for familiarization. The affective state of the user can indicate whether they should be interrupted or not (e.g., when they are in a state of deep concentration). Altogether, the foundation that this work lays could lead to improved human-machine interactions where people feel supported by machines in their creative tasks, without feeling like their creative agency is automated away.

4.8.4 Limitations

Three main limitations of our study should be noted. First, our study was based in a lab setting and took place in a limited amount of time. Typical usage of comparable machines will often take place in users familiar environments (e.g., home or makerspace) or at a workplace that is shared with others; either in the background, or as active collaborators. Furthermore, interactions with fabrication machines in naturalistic settings would also likely take much longer than the single-session tasks that were performed in our experiment. The often more complex naturalistic situations could call for different types of instrumentation, which was beyond the scope of the present study.

Second, the population that we studied consisted of nearly all novices. While this offered us opportunities to observe various breakdowns, it also constrained the level of interaction that participants undertook with the machines; they did not feel comfortable to experiment with various settings. Furthermore, expert users likely show different attention patterns, which could lead to capture changes (e.g., as their gaze might not follow their hands).

Third, our sensing setup was limited to cameras capturing the tabletop area. We did not evaluate the addition of a foot camera for the sewing machine, nor the addition of embedded sensors or API streams coming from the machines.

4.8.5 Future Work

There are several directions in which this research could be extended. First, studying expert users longitudinally could reveal habits, as well as a richer understanding of evolving human-machine collaboration patterns. Second, applying our instrumentation approach to other types of machines, such as those of a larger scale, could inform its generalizability beyond the focused domain of tabletop-based creative fabrication machines. Third, we envision future work to apply the findings of this chapter in order to instrument fabrication machines with proactive Wizard-of-Oz responses to interaction breakdowns, as indicated through implicit and explicit signals. Finally, we plan to implement automated processing of real-time sensing, in order to gradually work towards achieving machines that intervene based on situations and signals laid out in this chapter. While touched upon lightly in this chapter, we envision real-time sensing of pose and facial expression data (e.g., OpenPose/OpenFace) to detect signals in an automated manner and respond in real-time.

4.9 Conclusion

We investigated how tabletop-based fabrication machines should be instrumented to facilitate interaction analysis. We conducted a user study (N=12) where we captured multi-view camera and semi-structured qualitative interview data from participants performing three tasks on a pen plotter, clay extruder, and sewing machine. First, we found that multi-view camera instrumentation from both the participant's and the machine's perspectives tends to capture most relevant interaction signals, while enabling a lightweight instrumentation setup. We propose a protocol that interaction designers can use to determine the right instrumentation to capture interactions around tabletop-based fabrication machines. Second, we share FabriCam-5, our five-camera multi-view dataset of novice interactions with tabletop fabrication machines. Initial analyses of this dataset along with interview data from the study uncovered potential signals of interest for each machine across four dimensions: human state, machine state, artifact state, and contextual state.

This chapter lays the methodological groundwork for studying interactions around creative fabrication machines. The observation that implicit behavioral cues seem to precede explicit help-seeking, combined with the instrumentation methods developed here, motivated the Wizard-of-Oz study described next, in which we tested whether acting on those implicit cues through proactive assistance improves the making experience.

CHAPTER 5

PART III: PROTOTYPING INTERACTIONS AROUND MACHINES

Most fabrication machines demand explicit commands and provide explicit feedback, requiring the user's full attention. A 3D printer waits for a file, prints it, and beeps when done; a laser cutter requires precise settings before each cut; and a sewing machine follows the user's hands but offers no guidance about where those hands should go. This command-and-control paradigm works well for experts who know what they want and how to achieve it. For novices learning a craft, or for experienced makers attempting something new, the machine's silence can be unhelpful at best and frustrating or discouraging at worst. Mixed-initiative interaction [Horvitz, 1999] offers an alternative: systems where both human and machine can initiate action depending on the situation. The machine might notice confusion and offer help, which the user might accept or decline. Neither party has permanent control; instead, initiative flows to whoever is best positioned to act at each moment. For fabrication machines, this vision remains largely theoretical.

Chapter 4 observed times where implicit behavioral signals such as hesitation, gaze patterns, and postural shifts precede explicit requests for help, and that these signals can be captured through carefully positioned cameras. This chapter presents an exploratory Wizard-of-Oz study investigating how users respond to different levels of machine initiative during a physical fabrication task, asking three research questions:

1. **Subjective experience:** How does machine initiative level affect users' perceived timeliness, helpfulness, and annoyance of assistance?
2. **Task performance:** How does machine initiative level affect task duration

and verbal interaction patterns?

3. **Emergent interaction:** What patterns of human-machine interaction emerge during wizard-assisted fabrication, and what do these patterns reveal about the design space for collaborative machines?

We[1] focused on sewing, a loosely procedural activity performed on a machine that combines digital sensing possibilities with physical material constraints (see Section 5.3.2). We compared two assistance conditions: proactive (hereafter also "Active") and reactive (hereafter also "Passive"). In the reactive condition, the wizard intervened only when study participants explicitly asked for help or made errors that would prevent task completion. In the proactive condition the wizard additionally intervened when behavioral or contextual cues suggested confusion or impending difficulty: hesitation before a step, incorrect hand positioning, gaze patterns indicating uncertainty, or approach to common error points. Twenty participants, all self-identified beginners, completed a pillow-sewing task twice: once with reactive machine assistance, and once with proactive machine assistance. The wizard observed participants through four synchronized camera streams and delivered interventions through two modalities: a text-to-speech voice that appeared to come from the machine, and a projected dot that could point at specific locations on the workspace. The combination of verbal and spatial guidance ("this lever here") grounded instructions in physical locations. Through post-task questionnaires, semi-structured interviews, and analysis of video transcripts, we investigated how initiative level affected both subjective experience and observable behavior. The resulting CoSew-4 dataset, including synchronized multi-view video

[1] This work is intended for future submission to publication, likely as Bremers, A., Guimbretière, F., & Ju, W. (2026). Augmenting Fabrication Machines for Interaction Intelligence. Venue to be determined.

and transcripts from all sessions, is made available for further research.

The contributions of this chapter thus include: (1) findings that proactive assistance is perceived as significantly more timely, and likely more helpful, by novice users without increasing perceived annoyance; (2) evidence that subjective experience diverges from task performance, with strong learning effects dominating completion times while initiative level affects how users experience the interaction; (3) an approach for prototyping co-creative machine behaviors, including multi-camera observation, speech output, and spatial pointing; (4) qualitative insights on the social and emotional dimensions of human-machine fabrication, including the value of confirmation and encouragement alongside technical guidance; (5) design implications for fabrication machines that adapt initiative to user expertise and preserve human creative agency; and (6) a synchronized multi-view dataset, CoSew-4, which will be made available for future research upon acceptance of the connected publication.

The chapter proceeds as follows. We first review related work on machine initiative, co-creative systems, task guidance, and Wizard-of-Oz methodology. We then describe the recording setup and its design requirements. The study methodology section details the sewing task, experimental conditions, wizard protocols, and data collection procedures. The analysis section presents statistical results from Likert scales and behavioral measures alongside qualitative themes from video transcripts and interview data. We conclude with a discussion of implications for designing collaborative fabrication machines, limitations of the current study, and directions for future work.

5.1 Designing Machine Behaviors for Collaboration

The theoretical foundations established in Chapter 2 frame aspects of making that co-creative machines should preserve and support. This section reviews prior work along the three dimensions of initiative, co-creativity, and task guidance, and discusses the applicability of the Wizard-of-Oz method to prototyping machine behaviors.

5.1.1 Initiative

Building on the theoretical foundations established in Chapter 2, this section focuses on how initiative has been operationalized in empirical studies of physical collaboration. Mok et al. [2015] studied initiative in a physical collaboration context, using wizard-controlled robotic drawers that either anticipated user needs (proactive) or responded to gestures (reactive). They found that while expressive robots were experienced as more engaged, proactivity could negatively affect participants' perception of their social status relative to the robot. This suggests that initiative in physical collaboration contexts may carry social-relational meanings beyond mere functionality. Flemisch et al. [2016] proposed a framework distinguishing levels of human-machine cooperation, from shared control of immediate actions (the "sharp end") to cooperation on higher-level goals and navigation (the "blunt shaft").

This layered view suggests that initiative can operate at multiple levels simultaneously: a fabrication machine might execute low-level motions autonomously while deferring to human judgment on creative decisions.

5.1.2 Co-Creativity

As reviewed in Chapter 2, a growing body of work has explored co-creative systems that collaborate with humans on artistic and making tasks. Physical co-creative robots can be more satisfying than screen-based agents in motivating creative exploration [Lin et al., 2020b], and collaborative drawing systems can assist rather than replace human creativity through roles such as corrective drawing, predictive drawing, and scene completion, although artists remain skeptical of automation in creative work [Jansen and Sklar, 2021]. In fabrication specifically, design fictions have articulated that collaborative machines need accessibility, fluidity, and concurrency [Kim et al., 2017], and research into collaborative 3D printing has shown that real fabrication workflows rarely follow the linear design-then-fabricate model assumed by conventional pipelines [Goudswaard et al., 2024]. Xiong et al. [2023] outlined various classifications for human-machine collaboration in additive manufacturing, including distinctions between active and supportive human-machine relationships. This classification is relevant to our experimental design, where we compare a machine that actively offers assistance (proactive) with one that provides support only when asked (reactive).

These works collectively suggest that collaborative fabrication machines would need to: sense user state and task progress, select appropriate moments for intervention, communicate intentions legibly, and support rather than override human creative agency. However, testing these capabilities requires methods for simulating machine intelligence before it is technically implemented.

5.1.3 Task Guidance

Our study resembles task guidance research in some respects, though with an important distinction: study participants were told the machine was "a partner that can help," not a guide to follow. Traditional task guidance systems provide pre-loaded procedural information without environmental sensing; the worker supplies situational awareness [Ockerman and Pritchett, 2000, Sheridan, 1992]. Modern systems increasingly incorporate sensing and audio interfaces that can deliver guidance without disrupting visual attention [Syberfeldt et al., 2015, Guarese et al., 2024, Zue and Glass, 2000].

Our wizard-operated system bridges these approaches: it senses user state (through the wizard's observation) and delivers guidance through speech and projected pointing, but the guidance content responds to the situation rather than following a fixed procedure. Prior work has evaluated such systems through task performance measures, workload assessments, and Likert-scale ratings including helpfulness, pleasantness, and intrusiveness [Wu et al., 2020, Begault et al., 1996, Guarese et al., 2024].

5.1.4 Applying Wizard-of-Oz

Chapter 4 introduced Wizard-of-Oz methods for eliciting and capturing interactions. Here we focus on WOz as a method for prototyping machine behaviors, simulating how a system would act in order to study user responses.

The WOz method offers several advantages for interaction research. First, it enables rapid prototyping of robotic behaviors without requiring fully developed autonomous systems, allowing researchers to quickly iterate on designs and gather user feedback before committing significant development resources. Second, it facilitates human-centered design by enabling researchers to gather authentic human feedback on robot behaviors in realistic settings, ensuring that the eventual system is user-friendly. Third, it provides cost-efficiency by simulating autonomy rather than building it, avoiding the high costs associated with developing and testing actual autonomous capabilities during the prototype phase.

However, WOz studies also present challenges. For one, Wizard-of-Oz studies often require careful planning and are frequently tailored to highly specific applications [Porcheron et al., 2021]. The wizard's skill is critical to the method's success; wizards cannot perfectly replicate the exact experience users will encounter with a fully autonomous system, potentially leading to fatigue, errors, or improvisations that do not reflect the intended system behavior. For extremely complex systems with unpredictable outputs, human operators may struggle to simulate responses convincingly and consistently.

For fabrication machines specifically, Kim et al. [2017] presented design fictions exploring the concept of Human-FabMachine Interaction, proposing that collaborative fabrication machines need accessibility, fluidity, and concurrency. Their work highlights the requirements of WOz studies in fabrication contexts, where physical material constraints and safety considerations add layers of complexity beyond typical HRI scenarios.

5.2 Wizarding Co-Creative Interaction

Unlike conventional WOz setups that simulate conversational agents or screen-based interfaces, our approach addresses the unique requirements of physical making contexts: the need for real-time observation of material manipulation, spatially grounded communication, and intervention timing calibrated to the rhythms of craft work. This section describes the design requirements, system architecture, and wizard protocols that enabled the study of proactive versus reactive machine assistance.

5.2.1 Design Requirements

The setup design was guided by four requirements that emerged from prior design explorations (Chapter 3) and the instrumentation study (Chapter 4):

- Observation without intrusion. The wizard must observe the study participant's actions, attention, and affective state in sufficient detail to make informed intervention decisions, yet the observation infrastructure cannot interfere with natural making behavior. Prior work established that egocentric and machine-mounted camera perspectives provide comprehensive coverage of fabrication interactions without requiring cameras that participants must consciously avoid. We adopted this two-perspective approach as an observational foundation.
- Low-latency intervention. Opportunities for timely assistance during fabrication often occur within narrow temporal windows. A user hesitating

before lowering the presser foot, for example, presents an intervention opportunity that may last only one to two seconds before the user either proceeds (potentially making an error) or explicitly asks for help. The system architecture needed to support intervention latencies under two seconds from wizard decision to user-perceptible output.

- Spatially grounded communication. Fabrication assistance often requires spatial reference: "that lever," "the edge of the fabric," "here." Purely verbal assistance without spatial grounding forces users to map abstract descriptions onto their physical workspace, increasing cognitive load and potentially causing confusion.
- Flexible initiative levels. To study the effects of machine initiative, the platform needed to support cleanly separated experimental conditions. The same underlying infrastructure should enable both reactive assistance (responding only to explicit requests or observed errors) and proactive assistance (anticipating needs based on task state and user behavior) with consistent interaction modalities.

5.2.2 Instrumentation Design

The resulting setup facilitates observation, intervention, and condition management. The wizard observed participants through four synchronized video streams, each capturing different aspects of the interaction:

- *Egocentric view (participant-worn pinhole camera glasses):* This stream captured the participant's visual attention, revealing where they looked during decision points and showing hand movements from their own per-

Figure 5.1: Camera angles for the sewing study were chosen such that the study participant, the work piece and machine, the needle area and the handwheel were visible. (Left: front view, top right: eyeglass camera, middle right: side camera, lower right: workspace fisheye camera.)

spective. The pinhole form factor minimized awareness of being recorded compared to more obtrusive head-mounted cameras.

- *Machine-mounted 360° view (Insta360 camera):* Mounted on the sewing machine's frame, this camera captured the participant's face, upper body, and the workspace in a single continuous view. The participant-facing hemisphere provided facial expressions and body posture; the workspace-facing hemisphere showed material positioning and machine state.
- *Handwheel webcam:* A dedicated webcam focused on the sewing machine's handwheel, which is a control that novices frequently neglect but that experienced sewists use regularly. This stream allowed the wizard to monitor whether participants were using proper technique for starting and stopping stitches.

- *Fisheye workspace camera:* Mounted near the needle, this camera provided a detailed view of the sewing area, capturing fabric alignment, pin placement, and stitch formation. This view allowed for detecting subtle errors that might not be visible in wider shots.

The wizard accessed these streams through a combination of interfaces: the 360° and pinhole feeds via mobile device applications (Insta360 app and Ti2Cam respectively), and the webcam feeds via Open Broadcast Studio on the wizard's computer. While this multi-device arrangement introduced some complexity, it provided the flexibility to position each video source optimally on the wizard's workspace (see Figure H.1).

Our system provided two complementary output modalities:

- Speech output used text-to-speech synthesis through system speakers positioned to appear as though the machine itself was speaking. Pre-scripted utterances were triggered via keyboard hotkeys, allowing rapid intervention without the latency of typing. Scripts included greetings, task guidance at multiple levels of detail, encouragement, error corrections, and meta-conversational phrases ("Do you need help?", "Would you like me to call a human for assistance?"). Free text responses remained an option to facilitate a broad range of interaction.
- Spatial pointing used a projector to display a moveable white dot on the workspace table. The wizard controlled the dot's position through a pygame interface, enabling real-time pointing at specific locations such as the presser foot lever, the edge of fabric to be aligned, or the location where a pin should be placed.

Furthermore, the system accounted for two experimental conditions. Two Python scripts implemented the active (proactive) and passive (reactive) experimental conditions. Each script defined similar utterances but differed in the protocols governing when the wizard should trigger them. In the passive (reactive) condition, the wizard triggered interventions only in response to: (1) explicit verbal requests for help from the study participant, or (2) clearly observed errors that would prevent task completion (e.g., attempting to sew without lowering the presser foot, fabric bunching in a way that would jam the machine). The wizard did not intervene based on anticipated problems, hesitation, or suboptimal technique. In the active (proactive) condition, the wizard additionally triggered interventions based on anticipated needs inferred from task state and user behavior. Indicators that could prompt proactive intervention included: hesitation before a task step, incorrect hand positioning, gaze patterns suggesting confusion (e.g., looking around the machine as if searching for a control), or extended pauses that might indicate uncertainty. The proactive interventions were designed to feel anticipatory rather than intrusive by offering guidance just as a participant might be beginning to wonder what to do next.

5.3 User Study

We conducted a Wizard-of-Oz interaction elicitation study in order to collect multi-view video data on participants' interactions with proactive and reactive assistance coming from the machine. The wizard (the first author in all sessions) followed structured protocols for each condition while maintaining flexibility to respond to unexpected situations. In both conditions, if an initial intervention received no response, it was repeated at an increased level of detail. If participants remained stuck, the machine offered to "call a human for assistance," at which point the experimenter would enter to help directly. The distinction between conditions was not whether to intervene, but when: reactive interventions waited for explicit signals of difficulty; proactive interventions anticipated difficulty based on task state and user behavior (e.g., hesitation, incorrect hand positioning, gaze patterns suggesting confusion). The protocol was reviewed by Cornell University IRB under exempt number #IRB0148868.

5.3.1 Setup

Figure 5.2: In addition to an instruction sheet and camera glasses, the study participant was provided with all sewing materials: two pillows, pre-cut fabrics, a threaded sewing machine, scissors, a ruler, pens, chalk, glue, pins, eyes, and scrap fabric for practicing. *Note:* the sewing machine in the final study was outfitted with a fisheye camera not shown here, but depicted in Figure 5.1 and Figure H.1.

The study took place around a table that was set up with a Brother XM3700 sewing machine, materials, and augmentation/capture system (see Figure 5.2). As for materials, we provided participants with two examples of completed pillow plushies (each of which had the final seam in a different location), two pillow fillings, precut cotton quilting fabric in various colors, precut felt legs and mouths, plastic eyes, glue, scissors, a ruler, and extra wound bobbins. The table was instrumented with 4 cameras (egocentric camera through participant-worn pinhole camera glasses, a 360° machine-mounted Insta360 camera, a simple webcam to the right side focused on the handwheel of the sewing machine, and a custom-built fisheye camera mounted close to the machine needle to get a detailed view of the sewing area), a projector, and a speaker. A wizard was seated behind a screen in the room, and used a custom console to trigger audiovisual interactions that occur from the speakers and the projector in the room, which appeared to originate from the machine as first-person statements. The wizard was able to observe all camera streams in real-time[2]. Interventions were triggered either by explicit requests for help or observed slips, or at anticipated needs, depending on the condition.

The wizard console consisted of a Visual Studio project running one of two Python scripts, depending on whether the condition was Active or Passive. In each script, several sentences were preprogrammed to be triggered by a hotkey; if the wizard pressed this key, a female TTS voice would read the sentence in question (see Table 5.1 for an overview of all hotkey and sentence mappings for both conditions). In addition, a pygame window let the wizard control a white dot that was projected on the table in front of the sewing machine, which could point out things and ground the TTS help in physicality. Simultaneously, Open Broadcast Studio was used to provide a live view of the right side webcam and the fisheye camera and allowing the wizard to record both cameras in a synchronized manner. The wizard used an iPad running the Insta360 app to access a live view of the 360° machine-mounted camera, and an iPhone running the Ti2Cam app to access a live view of the pinhole camera glasses.

[2] Due to vendor app issues, the egocentric camera perspective would at times disconnect, but at these times the wizard would use the other camera perspectives for visual information.

5.3.2 Task Design

The task was to sew a throw pillow with eyes and limbs (i.e., a simplified plushie). Using a sewing machine, unlike the pen plotter and clay extruder, forces the user to contribute to the interaction, whereas a novice might otherwise be accustomed to observing without interacting (see Chapter 4). The task was originally designed to include multiple plushie shapes, but pilot testing revealed this exceeded novice capabilities within the time constraint. The final design used a single throw-pillow form, modeled after a YouTube tutorial [Tock Custom, 2022], making it suitable for novice sewists during a user study that was meant to be capped to 1 hour for two tasks. The task can be seen as a loosely procedural task [Eiriksdottir and Catrambone, 2011] with open-ended elements (e.g., fabric choice, limb placement) that give the user creative agency (preserving the values in Chapter 2). For customization, participants could choose the fabric to use and the color, shape, amount, and position of the mouth and legs. After the testing of the task by the first pilot participant, it was decided to pre-cut the fabric to the right size and with pre-cut holes for the eyes, to reduce the task complexity and keep the task to a reasonable time. The task and setup were piloted by four pilot participants (P1–P4) before the setup and task were finalized and the study began. Figure 5.3 shows the task outcome as it was designed. At the bottom, the initial plushie was made by the first author, and did not feature pre-cut fabric or holes. The two plushies above that were made during a first pilot with pre-cut fabric. For these, the fabric was cut to a slightly smaller size in order to maximize the number of usable pieces out of each fat quarter sized cut of quilting fabric. The two pilot plushies were displayed to the left of the participant during each trial to give participants a reference during the task.

Figure 5.3: Two iterations of sewing task designs. Bottom: initial design. Top two: a pre-cut and pre-punched design optimized for efficient fabric usage, completed by pilot P1 (who had prior sewing experience), and displayed to all study participants as a reference.

5.3.3 Procedure

At first, there was a familiarization task. In this task, participants were asked to sew a line on a piece of scrap fabric. During this task, the experimenter was present to help with setup, for safety reasons. The experimenter asked the participant if they had ever used a sewing machine before, and would then explain the functions of the presser foot lever, the foot pedal, and the handwheel, and instruct participants on how to hold the fabric and how to cut the thread after sewing (the cutter on the side of the machine was not introduced). Participants could then try for themselves to sew on the scrap fabric while the experimenter watched, and they had the opportunity to ask questions about machine operation. After the familiarization task was completed, the experimenter used a clap board to synchronize camera streams and from then on, the participant could start making the first plushie. For this task, they were given a sheet with instructions that listed five steps. These instructions were deliberately limited in information content, so that the participant would likely need help during the task. The steps were based around a modified pillowcase sewing tutorial to account for the addition of eyes, feet, and a mouth. The steps that were provided to participants were:

Instructions for sewing a pillow plushie

1. *Gather*: gather supplies, fold fabric (right sides together)
2. *Prepare*: insert felt accessories (facing inward), pin edges leaving an opening.
3. *Sew*: sew seams, pivot at corners, remove pins.
4. *Assemble*: turn right side out, attach eyes, insert filling.

5. *Close*: pin the opening, sew the final seam, glue on mouth.

They then completed the task twice. After each task, a brief questionnaire was conducted where participants were asked to rate the perceived clarity, helpfulness, level of annoyance, trustworthiness, and timeliness of the machine's help on 5-point Likert scales. Participants could write open-ended comments on the questionnaire. At the end, a final interview probed participants further about their experiences interacting with the machine, their experiences with sewing, and their views on collaborative tools and machines more broadly.

5.3.4 Wizard Interventions

The wizard (the first author in all sessions) followed structured protocols using a hotkey mapping (see Table 5.1) while maintaining flexibility to respond to unexpected situations through free text responses. The distinction between conditions was not whether to help or how helpfully to respond, but when to initiate assistance: reactive interventions waited for explicit signals of difficulty (reacting to an explicit ask for help or visible slip); proactive interventions were machine-led (the intervention was based on an anticipated need inferred from human, machine, and task state). Conditions were presented to participants in a counterbalanced order.

Several elements of the protocol were common to both conditions. First, if an initial intervention received no response or did not resolve the difficulty, the wizard repeated the intervention at an increased level of detail. For example, an initial prompt might say "Check that the presser foot is down." If the participant did not respond, the follow-up might add spatial pointing and say "This lever here, push it down before you start sewing." This escalation ensured that participants would not remain stuck while preserving the lighter-touch initial intervention. If participants remained stuck after escalated assistance, the machine offered to "call a human for assistance." If the participant confirmed, the experimenter entered to help directly. This handoff mechanism ensured participant wellbeing while preserving the experimental manipulation. Participants experienced the machine as the primary source of assistance until its capabilities were explicitly exceeded. The wizard aimed for consistency in tone, pacing, and instructional content across conditions. However, as the wizard was responding to the participant, this meant that the conversation could get carried away into unpredictable directions (as seen in Section 5.5.4).

5.3.5 Participants

After four pilots, we recruited 20 study participants via flyers, e-mail and social media posts, and snowball sampling. Participants provided informed consent and were then introduced to the task. Almost all participants were novices to sewing. All participants completed both tasks (depicted in Figure 5.4). As an incentive, and to further encourage feelings of ownership of the final product, participants were allowed to keep what they made during the study.

Table 5.1: Interventions mapped to keys in Active vs. Passive conditions.

Key	Active Condition	Passive Condition
left	Make sure to put the right sides together	You did not put the right sides together
right	The feet need to face inward	The feet are facing the wrong way
g	The feet should be placed between the fabric	The feet are in the wrong position
down	Make sure not to add pins all the way	You forgot to leave an opening
return	Do not cut the thread too short	The thread was cut too short
backspace	Rest assured, I am here to help you	Now you need to rethread the needle
tab	Are you okay?	Are you okay?
space	Do you need help?	Do you need help?
s	Should I call a human for assistance?	Should I call a human for assistance?
w	Try to sew in a straight line	Try to sew in a straight line
e	The washer for the eye should be on the inside	The washer for the eye should be on the inside
r	Hello! I am your smart sewing machine.	Hello! I am your smart sewing machine.
t	What are we making today?	What are we making today?
y	This is my working area.	This is my working area.
u	Goodbye!	Goodbye!
i	That is right.	That is right.
o	Uh-oh.	Uh-oh.
p	I believe in you!	I believe in you!
a	Yes.	Yes.
z	No.	No.
x	You seem to have a sheet with instructions.	You seem to have a sheet with instructions.
caps lock	Make sure to push the filling inward before pinning the final edge	The filling should have been pushed inward before pinning the final edge
shift (L)	You are forgetting a step	You forgot a step
shift (R)	You are doing it correctly	You did it correctly

5.3.6 Data Processing

After the experiment concluded, we trimmed all video streams using the clapboard signal as a start[3]. From the 360° machine-mounted camera, a participant-facing view was extracted using Insta360 studio (at the UltraWide setting with aspect ratio = 1:1, pan = 177.3°, tilt = -28.3°, roll = 0.0°, FOV = 68°, distortion = 1.00, resolution = 800×800, bitrate = 10). The original 360° video data was saved

[3] For P16, Task 2, there was no clapboard signal. Instead, the machine's initial greeting was used as a synchronization point.

for potential future reference (e.g., to see detailed manipulation as part of re-threading the machine). The resulting dataset was processed into a four-camera view of all camera angles. We ran OpenPose and OpenFace on the dataset. Interview data and experiment audiostream data were processed using ffmpeg and transcribed using Whisper [OpenAI, 2022] on a local GPU, with manual transcript corrections done by the author using Aegisub [Aegisub Contributors, 2024]. The resulting transcripts were used to analyze moments of wizard intervention, as well as start- and end times of each task.

5.4 Analysis

In our experiment, we collected a synchronized multi-view video dataset consisting of four camera angles (egocentric pinhole glasses, 360° machine-mounted camera, right-side webcam and a fisheye workspace camera), post-task questionnaires with 5 Likert scales and open-ended comments, and post-experiment semi-structured qualitative interview data about participants' experiences with the machine. This section describes our analytical approach to each data source.

5.4.1 Quantitative Analysis

In this section, we cover the analysis approach we took for the Likert Scale questionnaires, as well as quantitative task duration and verbal interaction metrics.

Figure 5.4: The products study participants created during two sewing tasks (top = T1, bottom = T2).

5.4.1.1 Likert Scale Analysis

At the end of every task, participants rated the machine's assistance on five dimensions: clarity, helpfulness, annoyance, trustworthiness, and timeliness. All items used 5-point scales (1=Strongly Disagree, 5=Strongly Agree), responding to statements of the form "The machine help was [clear/helpful/annoying/trustworthy/timely]."

We treated Likert data as ordinal and applied Wilcoxon signed-rank tests to compare Active and Passive conditions. To address multiple comparisons across five related items, we applied Holm-Bonferroni correction and report both uncorrected and corrected significance levels. We also computed effect sizes (r) to assess practical significance independent of statistical significance. Given the counterbalanced design, we conducted mixed-design ANOVAs to test for Condition $\times$ Order interactions that might indicate task-order confounds.

5.4.1.2 Task Duration and Verbal Interaction (Word Count)

Audio streams were extracted from each task and transcribed using Whisper [OpenAI, 2022] with manual correction in Aegisub. For most videos the best audiostream came from the eyeglass camera; the 360° camera audio stream was used as a backup. We identified task boundaries using the machine's greeting ("Hello, I am your smart sewing machine") as start time and farewell ("Goodbye") as end time[4].

[4] For P11 T2 there was no "Goodbye" at the end, instead, the last utterance ("Okay") on the transcript before camera shutdown was used as the end time

From the transcripts, we computed three behavioral measures: task duration in seconds, participant word count, and machine word count. Annotations of inaudible speech and non-verbal vocalizations (singing, humming) were excluded; verbal fillers ("Oh," "Ehm," "Aha") were included.

We analyzed these measures using mixed-design ANOVAs with Condition (Active vs. Passive) as a within-subjects factor and Order (Active-first vs. Passive-first) as a between-subjects factor. We computed effect sizes (η_p^2) and conducted simple effects analyses to decompose significant interactions.

5.4.2 Qualitative Analysis

Qualitative approaches that we took consist of analysis of the video transcripts for interaction patterns, analysis of the interview transcripts for subjective experience patterns, and analysis of post-task written questionnaire comments.

5.4.2.1 Video Transcript Analysis

The 40 task transcripts (20 participants $\times$ 2 tasks) were analyzed for interaction patterns. We identified moments where participants established the machine as an interlocutor, engaged in collaborative repair of misunderstandings, provided or received emotional support, and requested human assistance. Representative excerpts were selected to illustrate the range and richness of interactions.

5.4.2.2 Interview Analysis

Semi-structured interviews were conducted immediately following completion of both tasks. Interviews lasted an average of 6 minutes 31 seconds (*SD* = 3:08, range 2:33–13:13) and covered participants' overall impressions, reflections on machine interactions, suggestions for improvement, and attitudes toward machine initiative in fabrication contexts. The guiding questions for the interviews are given in Appendix G.

Interview recordings were transcribed using Whisper, and analyzed using thematic analysis. We identified recurring themes related to learning and skill development, the value of proactive intervention, confirmation and emotional support, expertise-dependent preferences, and suggestions for improvement.

5.4.2.3 Post-Task Comments

Fifteen of twenty participants provided written comments on the post-task questionnaires. These comments were analyzed alongside interview data to triangulate themes and identify specific moments that shaped participants' assessments of each condition.

5.5 Results

This section presents findings from the analyses described above. We organize results around our three research questions: how initiative affected subjective experience, effects on task performance, and emerging interaction patterns.

5.5.1 Subjective Experience

As seen in Figure 5.5 and Table 5.2, participants' ratings were generally positive in both the Active and Passive condition. Differences can be seen in the distributions for Helpfulness and Timeliness.

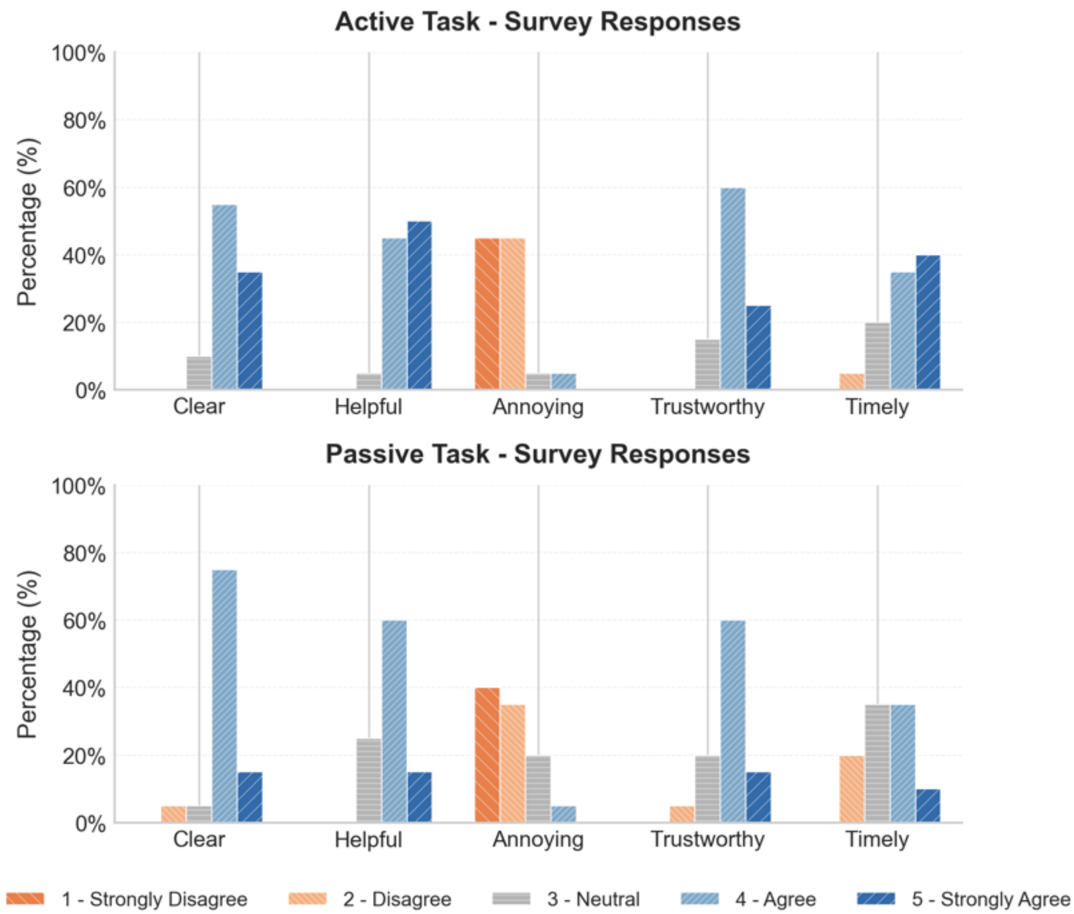


Figure 5.5: Likert scale responses to the statement "The machine help was Clear/Helpful/Annoying/Trustworthy/Timely."

Proactive assistance was also rated higher on helpfulness. This difference was significant before correction ($W = 12.0$, $p = .016$, $r = .65$), but did not reach the adjusted threshold after Holm-Bonferroni correction ($\alpha = .0125$). Active condition ratings for helpfulness were overwhelmingly positive (95% rated ≥ 4, with 50% at the highest rating) compared to perceived helpfulness in the Passive condition (75% rated ≥ 4, with 15% at the highest rating). The large effect size suggests practical importance.

Proactive assistance did not increase annoyance. Both conditions received low annoyance ratings with no significant difference between them (p = .498). No differences emerged for clarity (p = .258) or trustworthiness (p = .250). These findings indicate that proactive and reactive interventions felt similarly clear and trustworthy to novice users, and both types of interventions showed low annoyance ratings.

No significant Condition × Order interactions appeared for any Likert item (all $p > .40$), confirming that the observed effects are attributable to assistance condition rather than task order.

Table 5.2: Statistical comparison of Active vs. Passive conditions.

Dimension	***W***	***p***	**Effect Size (*r*)**	**Significance**
Timely	0.0	0.0010**	0.885 (large)	Significant**
Helpful	12.0	0.0164†	0.649 (large)	Trend†
Clear	12.0	0.2578	0.415 (medium)	n.s.
Trustworthy	6.0	0.2500	0.511 (large)	n.s.
Annoying	34.0	0.4978	0.223 (small)	n.s.

Note. N = 20 pairs. Wilcoxon signed-rank tests with Holm-Bonferroni correction for multiple comparisons. Effect size r: .10 = small, .30 = medium, .50 = large. ** $p < .05$ after Holm-Bonferroni correction. † $p < .05$ before correction; did not reach adjusted threshold (α = .0125). n.s. = not significant.

5.5.2 After Task Comments

Several participants explicitly valued the Active condition's capacity to anticipate and prevent errors. P5 wrote, "I love [that] the machine is proactive at times and gives me hints and prevents mistakes," while P18 noted after the Active task, "Much more proactive, really helped." P20, after experiencing the Passive condition, explicitly wished for proactive feedback: "For the help times, maybe it is also good to give me feedback before I ask."

Participants also commented on the learning aspect across tasks. P22 noted "It was easier because I had done it before," while P15 commented after the Passive condition (their second task), "Most of the time felt like I was alone, felt okay with that since I now mostly knew what to do." These self-reports are consistent with the significant Condition × Order interactions observed in the behavioral data.

One comment raised an important consideration for adaptive assistance design. P22 observed: "The machine was a bit more proactive. I could see that being annoying if you have experience." This concern aligns with prior work suggesting that proactive assistance may need to be calibrated to user expertise. Additionally, several comments touched on relational dimensions; P17 noted that in the Passive condition "It felt a little harder to talk to the machine," suggesting that machine initiative may influence the perceived relationship between user and machine.

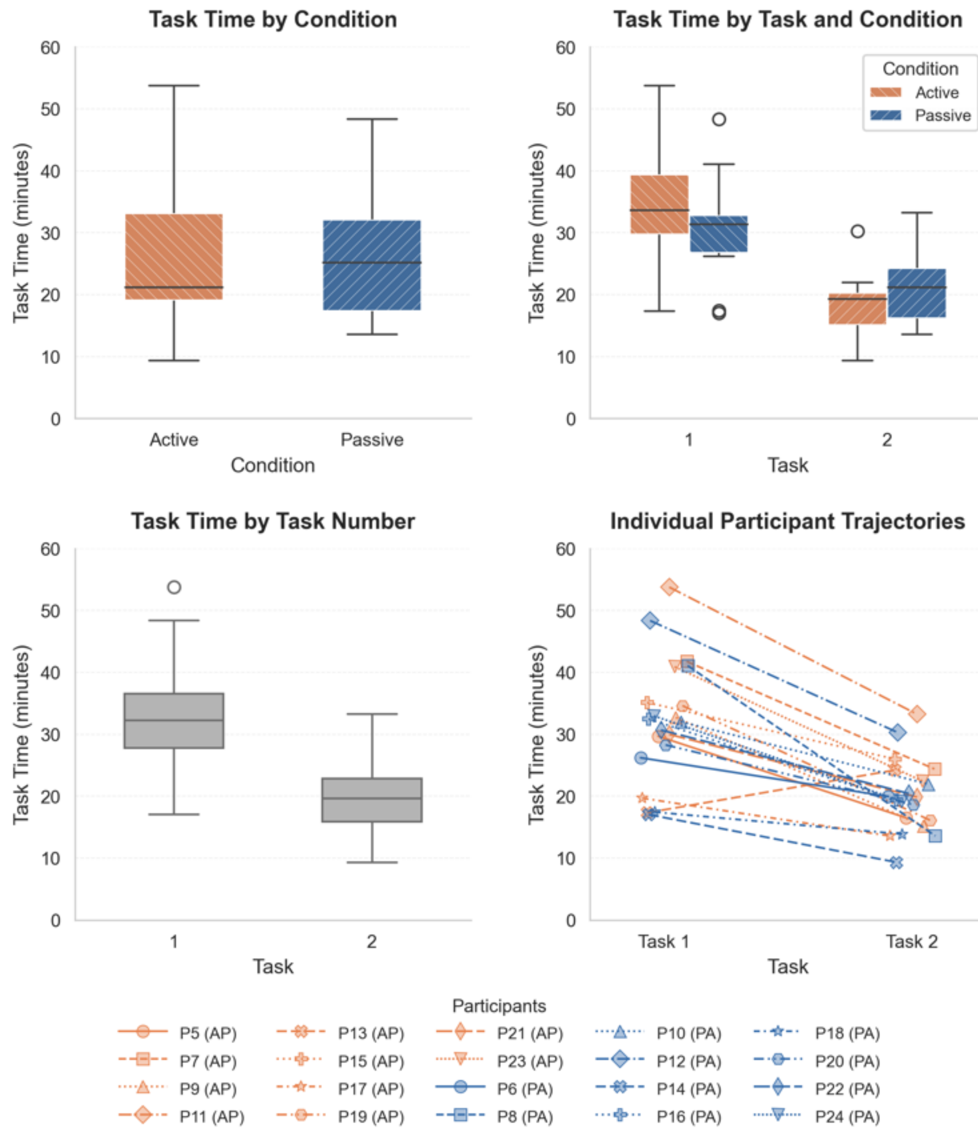


Figure 5.6: Task time distributions during the sewing study.

5.5.3 Task Performance

Task duration showed no main effect of assistance condition ($F(1,18) = 0.014$, $p = .907$) (see Figure 5.6, Table 5.3). Instead, a strong learning effect emerged: participants completed their second task faster than their first regardless of condition order. The Condition × Order interaction was highly significant ($F(1,18)$

= 51.74, $p < .001$, $\eta_p^2 = .742$). Simple effects analysis revealed that for AP participants (Active first), the Active task took significantly longer than the Passive task ($t(9) = 4.73$, $p = .001$), while for PA participants (Passive first), the pattern reversed, with the Passive task taking significantly longer ($t(9) = -5.58$, $p < .001$) (see Table 5.4). This crossover interaction indicates a strong learning effect: participants consistently performed faster on their second task regardless of condition. Between-groups comparisons at each time point revealed no significant differences between Active and Passive conditions for task duration (T1: t = 0.64, p = .529; T2: t = -0.95, p = .354) or participant word count (T1: t = 0.40, p = .692; T2: t = 0.57, p = .579). The significant Condition × Order interactions observed in the ANOVAs were attributable to a strong learning effect, with participants completing their second task approximately 12 minutes faster regardless of condition. The disconnect between subjective ratings and task performance is notable. Participants perceived the conditions as quite different even though this did not change completion times.

Table 5.3: Mixed ANOVA results for behavioral measures.

Measure	**Effect**	$F(1,18)$	p	η_p^2	**Significance**
Task Duration	Condition	0.014	.907	.001	n.s.
	Order	0.675	.422	.036	n.s.
	Interaction	51.740	<.001**	.742	Significant**
Participant Words	Condition	2.052	.169	.102	n.s.
	Order	0.000	.992	.000	n.s.
	Interaction	22.440	<.001**	.555	Significant**
Machine Words	Condition	44.021	<.001**	.710	Significant**
	Order	0.630	.438	.034	n.s.
	Interaction	69.858	<.001**	.795	Significant**

Note. N = 20 participants (10 per order). Mixed-design ANOVA with Condition (Active vs. Passive) as within-subjects factor and Order (AP vs. PA) as between-subjects factor. Effect size η_p^2: .01 = small, .06 = medium, .14 = large. **$p < .001$. n.s. = not significant.

The same pattern emerged for participant word count: no significant main effect of Condition ($F(1, 18) = 2.05$, $p = .169$, $\eta_p^2 = .102$), but a significant Condition × Order interaction ($F(1, 18) = 22.44$, $p < .001$, $\eta_p^2 = .555$) (see Figure 5.7). Participants spoke more during their first task (AP: Active $M = 757.4$ words vs. Passive $M = 259.1$ words; PA: Passive $M = 639.2$ words vs. Active $M = 372.3$ words), suggesting they asked more questions and verbalized more when unfamiliar with the task.

Machine word count showed both a significant main effect of Condition ($F(1, 18) = 44.02$, $p < .001$, $\eta_p^2 = .710$) and a significant Condition × Order interaction ($F(1, 18) = 69.86$, $p < .001$, $\eta_p^2 = .795$). The Active condition produced more machine speech at both T1 ($t = 3.44$, $p = .003$) and T2 ($t = 6.73$, p <.001), consistent with its proactive design. However, the magnitude of this difference varied by order: for AP participants, the Active condition produced substantially more machine output than Passive ($M = 497.6$ vs. $M = 88.7$ words), while for PA participants, the difference was smaller and reversed in direction ($M = 240.2$ vs. $M = 287.2$ words). This suggests that participants in their first task (regardless of condition) elicited more machine responses, either through proactive assistance (Active) or by asking more questions that triggered reactive responses (Passive).

Table 5.4: Simple effects: Active vs. Passive by Task (between-groups).

Measure	**Task**	***M* Active**	***M* Passive**	***t*(18)**	***p***	**Significance**
Task Duration (s)	T1	2013.0	1839.1	0.642	.529	n.s.
	T2	1119.5	1269.2	-0.951	.354	n.s.
Participant Words	T1	757.4	639.2	0.402	.692	n.s.
	T2	372.3	259.1	0.565	.579	n.s.
Machine Words	T1	497.6	287.2	3.443	.003**	Significant**
	T2	240.2	88.7	6.734	<.001**	Significant**

Note. Independent-samples t-tests comparing Active vs. Passive at each task level. At T1, AP group received Active and PA group received Passive; at T2, the reverse. ** $p < .01$. n.s. = not significant.

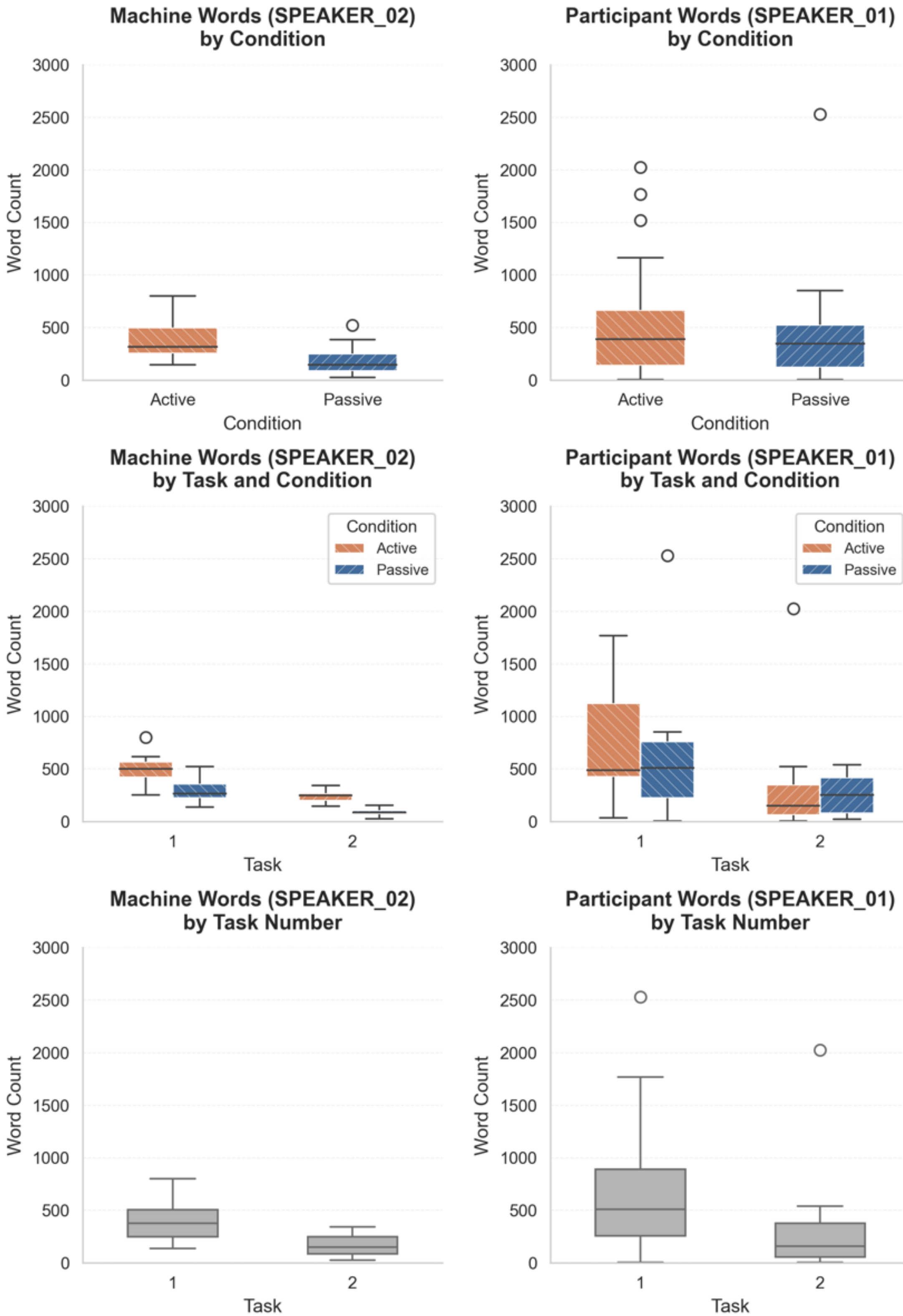


Figure 5.7: Word count distributions during the sewing study.

5.5.4 Interaction Patterns

Analysis of video transcripts, supported by interview data, revealed interaction patterns extending beyond task guidance.

5.5.4.1 Establishing the Interaction Rules

When the machine greeted participants, several were uncertain how to respond. P5 asked "Do I speak to it?" P10 asked the experimenter "Am I supposed to respond to it?" Once the machine confirmed its addressability, participants engaged readily. Some developed ongoing relationships: P6 named the machine "Marceline" and tested whether this name persisted across tasks, and P19 named it "Dhaga."

5.5.4.2 Collaborative Repair

The instruction "right sides together" caused repeated confusion. Participants interpreted "right" as direction or correctness rather than the printed fabric surface. Successful clarification required iterative exchanges, with participants demonstrating their interpretation and the machine correcting them. Language alone was insufficient; participants needed to act and receive feedback.

P5_T1 (00:01:05–00:02:40) gives us an example of successful repair through extended negotiation:

Machine: Make sure to put the right sides together.
P5: What's right?

Machine: The right sides should face inward.

P5: Like that?

Machine: Turn the fabric over.

P5: How about that?

Machine: Just the top one.

P5: Oh, eh. The top is the one with the holes?

Machine: Yes.

P5: So just like this?

Machine: Now do the same to the bottom one.

P5: Ah, so turn it inside out.

Machine: Yes.

P5: Great.

Machine: You are doing it correctly.

P5: Thank you.

From the interviews, it was also apparent that three specific aspects of the task generated consistent difficulty across participants.

As illustrated by the collaborative repair transcript above, the instruction to place fabric with "right sides together" confused participants who interpreted "right" as a direction or correctness rather than the printed surface. P14 captured the frustration: "I didn't realize that it was starting from an inside-out position on the fabric, so I just kept putting the fabric the wrong, like, the outside facing out, and the sewing machine's instructions did not clarify that until I'd been trying for a while."

The orientation of the pre-made legs also generated confusion. P18 noted: "The legs, trying to figure out where to insert the legs. And it is something that's also hard to explain with audio only." P24 agreed: "Figuring out how to put the legs in the middle, I feel like that was the most unclear direction."

The final seam, closing the pillow after inserting the filling, was universally identified as the most difficult step. P5 described the challenge: "The problem is: to fold both ends into inside and pin them down was pretty hard. And even if I pin them perfectly, once I start sewing somehow one of them will drift apart." P9 elaborated: "With the pillow, it was kind of hard to close the sides after I put the pillow in. It looks really bad. But yeah, that was the most frustrating part. I had to keep redoing it."

5.5.4.3 Emotional Support and Humor

Participants valued the machine's encouragement. "I believe in you" appeared at moments of uncertainty across sessions and was acknowledged warmly. P24 developed a characteristic "Why thank you" response (eight instances in Task 2). **P11_T1 (00:46:15–00:46:53)** displayed mutual expression of vulnerability as she decided to hand-sew the remainder of the task:

> *Machine*: I'm not offended that you won't use me.
> *P11*: You've been pretty helpful.
> *Machine*: I guess I'm clumsy.
> *P11*: No, I'm the real clumsy one.
> *Machine*: No! I believe in you.
> *P11*: Thank you. I think we're all doing great.

In interviews, participants consistently valued the machine's confirmatory and encouraging statements. This theme appeared across nearly all interviews, regardless of condition order:

> "The machine is very encouraging. It's like constantly giving positive confirmation. And I think it helps getting the basics down." (P11)

> "I think quite useful for me is the confirmation, because sometimes I'm in the uncertainty stage, and I truly need that kind of confirmation, even if that's wrong." (P20)

> "It's nice kind of getting the confirmation that I'm doing it correctly." (P24)

P11 offered a nuanced observation about the phrasing of check-ins: "I will say the phrasing like 'are you okay' makes me feel like I'm not okay. So like maybe, I don't know. 'Do you need help?' 'Is everything alright?'" This highlights that the specific word choice of proactive interventions can make a difference in how the same message is received.

Several participants noted that the encouragement reduced feelings of isolation during the task. P6 stated simply: "I liked talking to the machine because then I have company when I was doing this." P8 elaborated: "If they can be a collaborator, that would be great. Because especially when you are a beginner, you need somebody to learn with you. To make you feel less lonely."

Another observation was the successful use of humor as an interaction strategy by the machine. **P19_T1 (00:24:13–00:24:29)** contains an example:

P19: Sorry, I am taking so much time. I'm new to this. Thank you for the patience.

Machine: No worries. It's not like I go anywhere.

P19: That was a nice joke.

5.5.4.4 Learning Between Tasks

Task 2 transcripts were shorter than Task 1, indicating acquired competence (also see the word count analysis in Section 5.4.1.2). P24 explicitly acknowledged this transition in Task 2's first minute: "I just made one and now we're making take number two, so I think I know how to get started at least." The participant sometimes also referenced prior performance, as **P19_T2** opened with: "Let's do it a little bit better than last time." During the interviews, several participants explicitly commented on their learning between tasks:

> "I know exactly what to do for the second task. So I think the overall guidance or teaching process during the first task is effective enough such that I exactly know what to do for the second task." (P12)

> "The first time I had a lot of guidance from the voice agent because I don't really understand the process... Second time I don't need too much help because I know everything already. So I just replicate the steps." (P5)

> "I think since I hadn't been sewing for a while, the first time was a bit challenging but not too hard since it's just mostly sewing in straight lines. And the second time around it was pretty easy." (P7)

"I think the first one was not so good, but I think with the help and also with the experience from the first task, the task 2 is easier to do." (P21)

5.5.4.5 Human Handoff

When machine guidance proved insufficient, a consistent conversational sequence managed the transition to human assistance:

Machine: Do you need help?
Participant: Yes.
Machine: Should I call a human for assistance?
Participant: Yes, please.

The phrasing "call a human" positioned the machine as distinct from humans while treating human assistance as a resource it could mobilize. rethreading-related issues were the most common trigger.

These excerpts illustrate that human-machine interaction in the creative sewing context was rich, varied, and deeply social. Participants and machine engaged in personification, collaborative repair, bidirectional emotional support, and humor.

5.5.4.6 Proactive or Passive Guidance

Participants tended to describe the proactive assistance in positive terms, particularly when they had encountered it during their first task. P18, who received Passive assistance first and Active second, articulated the contrast:

> "The two interactions were different. [With] the first one I definitely felt the need for more proactive intervention. For example, if I was taking too long at a step or flipping something around too many times, it would have been great to have the intervention of the machine. And the second one was doing that, [and] was being a lot more proactive at a time where I already knew what I was doing, but still it was being proactive in a way that was helpful and was not too intrusive."

P20 described the ideal timing for intervention: "The reason I ask a question is I start to feel frustrated. If the machine or the system can notice that before I start to feel frustrated, just give me some suggestions. That is the most timely, I think." This observation connects to the quantitative finding (see Figure 5.5) that Active was rated significantly higher on timeliness.

P17, who received Active first, noted a qualitative difference: "I think I could tell there was a difference in how the machine was acting. The first one was talking more than the second time. So I thought it was more easier to talk to, I guess, in that sense."

However, several participants raised concerns about maintaining their own sense of agency when working with proactive systems. P20 described feeling constrained by early directive assistance:

> "One thing I really feel is not that okay [is] at the beginning. Because [there are] a lot of things I'm working on and the machine provides me directly, 'you should do that, you should do that,' and I feel like oh shit, the machine is annoying me... What you want is 'you think

> about this,' and the decision, the agency, is still [with] the user or the learner."

P13 offered a counterpoint, observing that the machine's physical limitations actually supported learning: "One thing [that] is good about it being AI and not human, is that if there was a human helping you, there's always this feeling that they might just intervene and do it themselves. Because if my mom was there or a teacher, they might just say, okay, let me do this tricky part. And now I had to do these tricky parts myself because the AI was limited in not having physical motor skills but just being able to talk."

P22 expressed mild discomfort with the social aspects of machine interaction: "I wouldn't be opposed, but it does feel a little creepy if it's too collaborative... It just feels like, it almost feels like it's not yourself anymore... And it felt weird a little bit sitting here, talking to a machine."

5.5.4.7 Requests for Visual Guidance

While participants generally appreciated the audio guidance, many expressed a desire for additional visual support:

> "I would say visual for a task like this would be better just because [the task] is something visual." (P7)

> "It would actually be cool if there was a panel, a video panel to see what I'm doing. [...] I think that would help me to understand what the machine is taking in." (P10)

> "Maybe pictures of how the things look like [for] each step so I can follow along better." (P17)

P12 specifically connected this to the pointing functionality: "I would wish the visual overlay to actively point at points of interest. Because I remember myself more or less having the stuff on the table flat and stationary... even if the cursor just kind of moves and moves through the area of interest, it will help a lot for me to understand."

However, P24 offered a contrasting view: "When I first looked at this, I was thinking, oh, potentially [I'd like] some visuals. But I kind of like that there are no visuals. It lets me focus on what I'm doing because there's a lot going on."

5.5.4.8 Expertise-Dependent Preferences for Initiative

A consistent theme across interviews was that preferences for machine initiative depend on user expertise. The participants, who all had limited experience, mentioned that they valued proactive assistance, while also recognizing that more experienced users might find it intrusive:

> "A guide for when you first start out, and then a collaborator for when you're more like an expert level and you just need someone to remind you of certain things you should do [that] you might forget." (P9)

> "If I know what I'm doing, probably not. But if I need help, if I don't know what to ask, then I think it would be helpful for the machine

> to [take initiative] because it probably knows more and it can help me." (P15)

> "I think again, as a newbie, yes. Afterwards, once you're repeating instructions for tasks that I've done before so I remember it, then I would tone it down to a more passive assistance. But proactive assistance is definitely super helpful when it's the first time." (P18)

P23 framed this in terms of skill levels: "If I become a master of the sewing stuff, I definitely need the tool just following my orders. But if I'm just a beginner, I definitely want [it to be] more instructive. Medium level: be more collaborative."

P24 articulated how this connects to the purpose of the activity: "If I'm doing something that I enjoy because I enjoy doing it, I want to have full control over the entire thing and I don't want to offload any aspects of it. Whereas for something that's kind of more introductory, it's kind of like coaching."

5.5.4.9 Transfer to Other Domains

When asked about other contexts where they would value machine assistance, participants frequently mentioned cooking:

> "I hope it can help me cooking. Because the timing and the order of putting food inside is really important. So I think it could help a lot to tell you what thing to put and if it is ready for the next ingredient to add." (P8)

> "I use Gemini [while cooking]. Yeah, because my hands are dirty, I don't want to touch my screen." (P15)

> "I could see this being really useful in cooking actually... a little camera and it tells me like, oh, do this, you're not doing it right, that's going to burn." (P24)

Other fabrication activities also emerged. P7 and P11 both mentioned soldering as a domain where guidance would be valuable. P16 suggested laser cutters: "Like for instance, a laser cut machine, not everyone has a lot of experience with using it. [...] You have to go through the training, and then once you get the training and you have to do it yourself, you kind of forget the steps."

5.5.4.10 Summary of Results

The interview results complement and extend the quantitative results. Participants' explicit recognition of learning effects aligns with the observed decrease in task duration across tasks. Their positive descriptions of proactive assistance support the significant timeliness difference favoring the Active condition from the Likert scale questionnaires. The consistent requests for expertise-adaptive behavior suggest that future systems might benefit from mechanisms to adjust initiative levels based on user familiarity with both the task and the system.

The interviews also revealed aspects of the experience not captured by Likert scales. Examples are the emerging emotional dimensions of fabrication work, including feelings of isolation, uncertainty, and satisfaction. Participants valued the machine not only for its instructional content but for its presence as a companion during a potentially frustrating learning process. This social dimension of human-machine fabrication interaction could be further investigated in future work.

5.6 CoSew-4 Dataset: Preparation for Future Analysis

Beyond the findings presented above, this study produced a multi-modal dataset that lays the groundwork for future quantitative analysis of user behavior during machine-assisted creative work. A full computational analysis of behavioral signals preceding wizard intervention is beyond the scope of this dissertation. However, the data preparation and labeling work completed here establishes the necessary foundation for such investigations.

5.6.1 Speech and Transcript Data

All wizard-participant verbal interactions were transcribed using Whisper and manually corrected by the author using Aegisub [Aegisub Contributors, 2024]. The resulting SubRip (SRT) files contain timestamped utterances with speaker labels distinguishing participant speech from machine utterances and the occasional experimenter utterance. Each machine utterance was further categorized according to its communicative function (e.g., proactive guidance, reactive guidance, confirmation, encouragement) based on manual review of the transcripts. This labeled speech data enables future analysis of the temporal relationship between wizard interventions and observable user behaviors.

5.6.2 Body Pose Data

OpenPose [Cao et al., 2018] was applied to the participant-facing video extracted from the 360° machine-mounted camera. This yielded frame-by-frame body keypoint estimates using the BODY_25 model. The raw JSON outputs were processed into structured CSV files containing data for upper-body keypoints including the head, shoulders, and wrists. These pose estimates capture behavioral signals such as movement speed, stillness, hand position, and postural changes that may precede moments of hesitation or confusion.

5.6.3 Towards Automated Hesitation Detection

The qualitative findings from Section 5.5.4 suggest that participants exhibited observable behavioral changes when experiencing confusion or uncertainty. The wizard learned to recognize and respond to these moments proactively. The quantitative data streams described above could be used to computationally characterize pre-intervention behavioral signatures. Preliminary pose data inspection suggests that hesitation moments may be associated with decreased hand movement velocity, increased stillness, and changes in head orientation. However, rigorously validating such a hesitation detection approach would require: (1) systematic annotation of ground-truth hesitation moments across the full dataset, grounded in behavioral science theory; (2) feature engineering to extract behaviorally meaningful (not just computationally observable) signals from the raw pose estimates; (3) statistical comparison of behavioral patterns in time windows preceding wizard interventions versus matched control periods; and (4) evaluation of whether detected patterns generalize across participants and task contexts. This analysis would be a substantial undertaking that extends beyond the scope of this dissertation.

5.6.4 Dataset Availability

The prepared dataset is preserved for future research and will be released on Harvard Dataverse once the follow-up study is complete. This includes synchronized video streams, corrected transcripts with utterance labels, and processed pose feature files. The dataset enables follow-up investigations into the computational modeling of user states during fabrication tasks. Such work represents a natural extension of the Wizard-of-Oz methodology employed here: having established what a knowledgeable human wizard attends to when deciding to intervene, future systems might detect these signals automatically.

The transition from human-in-the-loop wizarding to fully autonomous intervention selection remains a significant open challenge. This dissertation identifies and motivates this direction but does not attempt to solve it. The prepared data resources described in this section provide a starting point for researchers pursuing this work.

5.7 Discussion

This study investigated how machine initiative affects user experience during physical fabrication. Using Wizard-of-Oz methods, we compared proactive assistance that anticipated user needs with reactive assistance that responded only to explicit requests or observed errors. Proactive assistance was perceived as more timely, and trended toward higher helpfulness ratings (see Section 5.5.1), without increasing annoyance. This section discusses what these findings mean for our research questions, evaluates limitations, and proposes directions for designing collaborative fabrication machines.

5.7.1 Evaluation of Research Questions

RQ1: Subjective experience. Proactive assistance was perceived as significantly more timely than reactive assistance ($W = 0.0$, $p = .001$, $r = .89$), a large effect that survived Holm-Bonferroni correction. Helpfulness showed the same direction with a large effect size ($r = .65$), though it did not reach the adjusted significance threshold after correction. Additionally, proactive assistance did not increase perceived annoyance. These findings support the viability of anticipatory machine assistance for novice users.

The finding that timeliness showed the strongest effect aligns with theoretical predictions from mixed-initiative research. Horvitz [1999] emphasized the costs of poorly timed interventions; our results suggest that in fabrication contexts, the cost of delayed intervention may exceed the cost of proactive intervention. When participants struggled without realizing they needed help, they experienced frustration. When the machine anticipated their confusion, they felt supported.

RQ2: Task performance. Task duration and verbal interaction showed no main effects of assistance condition. Learning effects dominated: participants completed their second task faster than their first, regardless of condition order. Simple effects analyses confirmed no difference between Active and Passive conditions at either time point. While this finding indicates that participants were not faster, it should be noted that the task goal was not to complete the plushie as fast as possible, but rather, we aimed to give participants a natural incentive to do well on the task, as they were allowed to customize and keep their made objects. For creative or learning-oriented tasks, user experience measures may be more informative than efficiency measures. How users feel about an interaction matters independently of how quickly they complete it.

RQ3: Emergent interaction patterns. Qualitative analysis revealed that participants engaged with the machine socially in ways that extended beyond task requirements. They named it, thanked it for encouragement, offered reassurance when it expressed self-deprecation, and sought its aesthetic judgment on creative decisions. These behaviors suggest that participants oriented to the machine not merely as a tool but as an entity capable of social participation.

The collaborative repair sequences (Section 5.5.4.2) illustrated both the limitations and possibilities of verbal guidance for physical tasks. Language alone was often insufficient for spatial reasoning, but patient, iterative exchanges could eventually achieve understanding.

5.7.2 Anticipatory Assistance for Novices

The finding that proactive assistance did not increase annoyance contrasts with concerns that anticipatory systems might negatively influence participants' perception of their own status [Mok et al., 2015]. One factor that might explain this is that our participants were novices performing a challenging task, and that they recognized their need for assistance. One of our participants did comment on the phrasing of interventions, which they would have preferred to be less authoritative (see Section 5.5.4.6).

Participants explicitly connected proactive intervention to feeling supported. However, participants also anticipated that their preferences would change with expertise. P23 articulated a progression from instructive (for beginners) to collaborative (for intermediate users) to responsive (for experts). This suggests that fabrication machines should adapt initiative level to user expertise rather than maintaining a fixed assistance style.

5.7.3 Social and Emotional Dimensions

Emotional support emerged as particularly valuable. Confirmatory statements ("You're doing great") reduced uncertainty. Encouragement ("I believe in you") sustained motivation through difficult moments. Several participants noted that the machine provided companionship during what could otherwise be an isolating learning experience.

The machine's inability to physically intervene, which might seem like a limitation, was perceived positively by some participants. P13 observed that unlike a human helper who might take over during difficult moments, the machine's lack of motor capability ensured that participants had to perform challenging steps themselves. The boundaries of machine capability served pedagogical functions by preserving opportunities for skill development.

These findings align with prior work on social responses to media [Nass et al., 1994] and suggest that fabrication machines should be designed with attention to relational dimensions, not just functional guidance. The social framing of machine assistance affected how users experienced the interaction.

5.7.4 Design Implications

Several principles emerge for designing collaborative fabrication machines:

Proactive assistance is viable for novices. Machines that can detect hesitation, confusion, or error-prone states should offer assistance before users explicitly request it. Waiting for explicit requests disadvantages users who do not know what questions to ask.

Initiative should adapt to expertise. Novices valued proactive guidance; they also anticipated it would become annoying with experience. Mechanisms to assess and adapt to user expertise would allow appropriate initiative levels throughout the learning trajectory.

Confirmation matters alongside correction. Participants valued being told they were doing well, not only when they made a mistake. Fabrication machines should communicate success as well as failure.

Verbal guidance has limits. Some challenges, particularly those involving spatial reasoning about three-dimensional transformations (orienting legs to face inward, closing a stuffed pillow), proved difficult to convey through language. These moments call for richer modalities, physical demonstration, or explicit handoff to human assistance.

Spatial grounding helps. The projected pointing mechanism, even though it was limited in its capability, allowed spatial reference ("this lever here") that grounded verbal instructions in physical locations. Fabrication guidance systems should connect language to space.

Social framing affects experience. Presenting the machine as a partner that offers support, rather than an authority that instructs, could influence user experience. Design choices about voice, phrasing, and interaction style shape the perceived relationship between user and machine.

5.7.5 Limitations

Several constraints limit generalizability of these findings.

All sessions were conducted by one experimenter acting as wizard. Structured protocols minimized variability, but different wizards might interpret intervention criteria differently. The wizard cannot perfectly simulate automated behavior; occasional delays when typing novel utterances may have affected the user experience.

The multi-device observation setup, while providing good visual coverage, occasionally suffered from connectivity issues with the mobile application of the pinhole camera feed. When this occurred, the wizard relied on the remaining camera perspectives, which provided sufficient information for most intervention decisions but occasionally missed gaze direction cues.

Participants were self-identified beginners. Findings regarding proactive assistance may not extend to experienced users, who might find anticipatory guidance intrusive. The participant sample was drawn primarily from a university community and may not represent broader populations of makers.

Sewing represents one fabrication domain with particular characteristics:

loose procedurality, material constraints, moderate safety concerns. Other domains involve different constraints that might change the dynamics of initiative. A CNC mill or laser cutter presents different hazards and requires different intervention strategies.

The study occurred in a controlled laboratory setting. Real fabrication contexts involve different social dynamics, tool availability, project complexity, and time constraints. Longitudinal use might reveal effects not apparent in single-session studies.

5.7.6 Future Work

The wizard learned to recognize behavioral patterns signaling confusion or hesitation: extended pauses, gaze patterns suggesting search, incorrect hand positioning approaching an error. The multi-view video dataset collected here provides a foundation for computational models that could detect these states automatically. Future work could annotate ground-truth hesitation moments and evaluate whether pose, gaze, and speech features could be used to predict intervention opportunities.

Our study compared fixed initiative levels. Future systems might adjust dynamically based on detected expertise, task phase, or explicit user preferences. P23's suggested progression from instructive to collaborative to responsive could be a trajectory that adaptive systems could support.

Longitudinal studies would clarify how preferences evolve as users gain experience. At what point does proactive assistance become unwanted? What signals this transition? How do users' mental models of machine capability develop over repeated interactions?

Extending to other fabrication domains would test generalizability. 3D printing, laser cutting, and CNC machining each involve different constraints, hazards, and user populations. The principles identified here may require domain-specific adaptation.

Participants requested richer visual guidance for spatially complex tasks. Future work should investigate how projected annotations, augmented reality overlays, or physical demonstrations complement verbal assistance. Which combination of modalities is most effective for fabrication guidance remains an open question.

5.8 Conclusion

This chapter presented an exploratory Wizard-of-Oz study investigating how users respond to different levels of machine initiative during a physical sewing task. Proactive machine assistance was perceived as significantly more timely ($W = 0.0$, $p = .001$, $r = .89$, surviving Holm-Bonferroni correction) and tended to be rated more helpful (not statistically confirmed after correction) than reactive assistance by novice users, without increasing perceived annoyance. These subjective benefits occurred despite no measurable differences in task performance, in which strong learning effects dominated. Participants engaged socially with the machine, valued emotional support alongside technical guidance, and anticipated that their preferences would change as expertise developed.

These findings support the viability of anticipatory fabrication machines while highlighting the need for expertise-adaptive initiative. The wizard's ability to recognize and respond to user states suggests that similar capabilities, implemented computationally, could enable fabrication machines that genuinely support human making rather than merely executing commands. The challenge ahead is building the sensing and intelligence to achieve what the wizard demonstrated: knowing when to help, how to help, and when to step back. We contributed the CoSew-4 dataset to enable future research.

CHAPTER 6

GENERAL DISCUSSION

This dissertation has explored the foundations of mixed-initiative interaction design for co-creative fabrication machines. The central question throughout has been: how can intelligent machines support human creativity in making without automating away the qualities that make craft valuable? Through theoretical grounding, systematic instrumentation, and experimentation with novices, this work establishes foundational principles for fabrication machines that act as attentive collaborators rather than autonomous executors.

6.1 Summary of Contributions

This dissertation makes contributions in three areas: methodological foundations for studying co-creative fabrication (Chapter 4), empirical findings on proactive machine assistance (Chapter 5), and design principles to balance creative agency with automation (Chapters 3 and 5).

Chapter 2 grounded this research in three larger theoretical traditions: Sennett [2008]'s analysis of craftsmanship, which establishes why the process of making matters independently of its products; Schön [2017]'s concept of reflection-in-action, which describes how practitioners learn through a "conversation with materials"; and Horvitz [1999]'s principles for mixed-initiative interaction, extended by Ju and Leifer [2008]'s framework for implicit interaction, which provide foundations for systems where humans and machines share control. These traditions revealed a tension at the heart of co-creative fabrication:

automation risks undermining the skill development that makes craft meaningful, yet fabrication machines offer capabilities beyond what manual work alone can achieve.

Interviews with maker entrepreneurs (Chapter 3) confirmed this tension empirically. Makers described the "joy of making" as central to their motivation and valued embodied engagement, skill development, and the satisfaction of seeing a project through. The design explorations in Chapter 3 then illustrated the design space through three relational frames: the machine as guide, as companion, and as adaptive collaborator.

Chapter 4 addressed a methodological prerequisite: before designing machines that take initiative, researchers need practical methods for observing the rich, multi-modal interactions that occur around fabrication machines. Through a study of 12 participants interacting with three tabletop machines (clay extruder, pen plotter, sewing machine), this work demonstrated that a minimal combination of egocentric and machine-mounted 360° cameras captures the majority of relevant interaction information, and proposed a generalizable protocol for instrumenting fabrication machines for interaction analysis. We noticed that implicit behavioral cues (hesitation, postural shifts, gaze changes) tended to precede explicit help-seeking, suggesting that a machine attending to these signals could intervene before a maker reaches the point of frustration. The FabriCam-5 dataset provides synchronized multi-view video to enable further research.

Chapter 5 compared proactive and reactive assistance during a pillow-sewing task with 20 novice study participants. Proactive assistance was perceived as significantly more timely than reactive assistance, with a large effect size that survived correction for multiple comparisons ($W = 0.0$, $p = .001$, $r = .89$).

Proactive assistance tended to be rated more helpful (not statistically confirmed after correction, $r = .65$), without increasing perceived annoyance. These subjective differences appeared despite no measurable effects on task performance, which was dominated by learning effects. Qualitative analysis revealed that participants engaged with the machine socially: they named it, thanked it for encouragement, and sought its aesthetic judgment on creative decisions. Emotional support emerged as particularly valuable. Participants also anticipated that their preferences would change with expertise, suggesting that fabrication machines should adapt initiative level rather than maintaining a fixed assistance style. The CoSew-4 dataset is available for future research.

6.2 Envisioning Experts Against Novice Evidence

A reader of this dissertation reaches Chapter 5 with a particular set of expectations. The dissertation opens with three makers (a crocheter, a potter, and a 3D printing expert) describing what they want from machines, and the introduction frames the dissertation around their tension between assistance and agency. Chapter 3 deepens this picture through interviews with maker entrepreneurs (gem cutters, jewelers, ceramicists, often with years of practice and an explicit identity as makers) and observations of sheet-metal operators with up to thirty years of experience on Amada bending machines. These are practitioners whose relationship with their tools is dense with tacit knowledge and whose concerns about automation, in line with Sennett [2008], are concrete and lived.

The empirical studies in Chapters 4 and 5 then turn to novices drawn from a university population. The 12 fabrication participants in Chapter 4 worked across three machines they had limited prior experience with, and the 20 sewing participants in Chapter 5 self-identified as beginners. The argument that proactive assistance was perceived as significantly more timely was demonstrated for this novice population. The gap between the framing population and the studied population is real and warrants direct address.

Three pragmatic forces drove this scoping on novices. First, experts in a given fabrication domain are unlikely to spend an hour completing a deliberately under-specified pillow-sewing task in a laboratory; experienced sewists, in particular, would complete the task quickly enough that the experimental manipulation would have little surface to act on. Redesigning the task to feature higher levels of complexity could address this issue, but would introduce other factors to the task that would hinder controllability of the user study and interaction elicitation effort. Second, the wizard's interventions, framed as a "helpful machine," were calibrated to communicate basic technique and machine operation, and an expert sewist receiving "check that the presser foot is down" would be a methodological mismatch rather than a meaningful test. Third, novices, aside from being the main population that was within reach for this user study, are themselves a significant population for collaborative fabrication: makerspaces, library fabrication labs, and consumer-grade machines (Cricut cutters, Bambu Lab printers, modern sewing machines) increasingly serve users whose first encounter with a machine is also their first encounter with the domain. Studying novices therefore represents a valuable scoping decision: this population is one for whom proactive assistance is plausibly useful.

As such, our studies provide evidence for one part of the vision and partial evidence for the rest. We show that proactive, well-timed assistance can be welcomed rather than intrusive in physical fabrication contexts, contradicting the worry, drawn partly from Mok et al. [2015]'s findings on robot proactivity affecting perceived status, that anticipatory machines could impose. We show that implicit signals (hesitation, gaze) at times precede explicit help-seeking, validating the design assumption underlying Ju and Leifer [2008]'s framework of implicit interaction as applied to physical making. We show that subjective experience and task efficiency can diverge, supporting Sennett's claim that the value of making is not reducible to the value of what is made.

What our studies do not show is what experts would experience. The makers of Chapters 1 and 3 described a relationship to tools that included frustration with constraints, pride in difficult work, and a desire to intervene in the machine's process. A proactive sewing-machine assistant that says "check that the presser foot is down" to a participant in their first hour is doing something different from what such an assistant would do for an experienced sewist working on a complex garment. The Chapter 5 finding that participants themselves anticipated a trajectory from instructive (P23: for novices) through collaborative (intermediate) to responsive (experts) is an explicit acknowledgment of this gap, voiced from within the study itself.

The risk here would be in generalizing from "novices welcomed proactive assistance" to "fabrication machines should be proactive by default." The findings do not support that generalization. The risk is also generalizing from one fabrication domain (loosely procedural sewing, with moderate safety concerns) to all fabrication; CNC machining, casting, glassblowing, and large-scale industrial fabrication carry constraints (safety, irreversibility, equipment cost) that may shift the cost-benefit balance Horvitz [1999] described. What the gap does not threaten is the methodological and theoretical contributions: the instrumentation protocol of Chapter 4, the datasets, the qualitative observations about implicit signals, and the design principles articulated below.

Closing this gap requires four lines of follow-on work: studies with experts in their own domains, longitudinal studies tracking individual users across the development of expertise, field studies in real fabrication contexts, and computational systems deployable at scales the wizard cannot reach. These are not independent: computational systems enable field studies, field studies clarify what experts actually do, and longitudinal data informs how systems should adapt. Section 6.6 below develops each in turn.

6.3 Implications for Design

The principles below are stated as insights about the novice population studied in this dissertation. Where they are most likely to extend to experts is noted where relevant.

6.3.1 Timing Matters As Much As Content

Chapter 5's comparison of proactive and reactive assistance held content similar while varying proactiveness, which influenced the perceived timeliness of the intervention. This finding aligns with Horvitz [1999]'s note on the costs of poorly timed interventions and extends it to physical making contexts. It also operationalizes the difference between foreground and background interaction that Ju and Leifer [2008] described: the proactive condition kept attentional demand low by intervening before study participants had to formulate a question or make a major mistake. For fabrication machines, the cost of delayed intervention may exceed the cost of proactive intervention, at least for novice users who do not know what questions to ask. As vision-language-action (VLA) systems become capable of producing a near-continuous stream of plausible suggestions, this finding inverts a common framing: the design problem is not how to generate enough useful content, but how to time a smaller number of useful interventions correctly.

6.3.2 Implicit Signals May Precede Explicit Requests

Chapter 4 observed that implicit signals (hesitation, gaze patterns suggesting confusion, expressive movements when encountering unexpected situations) often appeared before study participants explicitly requested help. Chapter 5 demonstrated that acting on these signals improved user experience without increasing annoyance. Together, these findings suggest that fabrication machines should attend to implicit as well as explicit user communication. Solely measuring response to explicit requests may underestimate the value of proactive systems by counting only the interactions that did happen rather than the difficulties that proactive intervention prevented.

6.3.3 Subjective Experience Diverges from Objective Efficiency

In Chapter 5, study participants perceived the conditions as quite different even though this did not manifest in different completion times. Learning effects dominated task completion times, while initiative level shaped subjective experience. This divergence suggests that efficiency metrics alone are insufficient for evaluating co-creative systems, since how users feel about an interaction matters independently of how quickly or well they complete it. This echoes the theoretical foundations: Sennett's craftsman values the process of making, not just the product. Schaldenbrand et al. [2024]'s CoFRIDA, discussed in Chapter 3, won Best Paper at ICRA 2024 for human-robot collaborative drawing, but focuses on what is drawn rather than the timing and appropriateness of machine initiative. The same observation applies more broadly: the VLA literature emphasizes task success (grasping, manipulation, instruction-following) over the relational and temporal dimensions of collaboration that this dissertation identified as central to user experience. Task-completion benchmarks may be especially misleading for creative work, where users' subjective experience can diverge sharply from objective efficiency.

6.3.4 Human-Machine Interaction Is Social And Emotional

Study participants in Chapter 5 engaged socially with the machine, valued emotional support alongside technical guidance, and treated the machine as an entity capable of social participation. This social engagement was not immediate;

it required an onboarding moment at the outset of the first interaction where participants established the rules of engagement (P5 asking "Do I speak to it?"). Once that frame was established, social interaction flowed naturally. This finding suggests that fabrication machines should be designed with attention to relational dimensions, not just functional guidance, and that the initial moments of interaction lay the foundation for the interaction that follows.

6.3.5 Machine Initiative Should Adapt to User Expertise

Novice participants valued proactive guidance but anticipated that the same guidance would become unwanted with experience. This is the most direct internal acknowledgment of the gap between this dissertation's framing and its evidence: even within a study of novices, participants pointed beyond their own present condition toward an expert future in which they would want different things. The trajectory implied is one of intensive support early that users internalize and apply independently later. Fabrication machines should assess and adapt to user expertise, providing appropriate initiative levels throughout the learning trajectory.

6.3.6 Preserve Human Agency in Creative Decisions

Throughout both studies, participants valued maintaining control over creative decisions while welcoming support with technical execution. The machine's inability to physically intervene was perceived positively by some participants because it ensured they had to perform challenging steps themselves. In Chapter 5, P13 observed that unlike a human helper who might take over during difficult moments, the machine's lack of motor capability meant the participant had to persist on their own. This suggests that the constraints of machine capability can function as design features rather than limitations: a machine that cannot take over preserves the conditions under which skill develops. As VLA systems become physically capable of more, this finding becomes a design choice rather than a happy accident: the question of what the machine should do, not just what it can do, becomes central. Sennett's concern with automation that replaces engagement becomes more relevant as machine capability grows.

6.4 Limitations

Section 6.2 discussed the most substantive limitation of this dissertation, the scoping to novice participants, in detail. Two further limitations warrant being mentioned briefly.

The Wizard-of-Oz method itself introduces methodological constraints. All sessions were conducted by the first author, who acted as wizard, designed the interventions, and had deep knowledge of the task. Structured protocols minimized variability, but different wizards might interpret intervention criteria differently. The wizard's expertise may also be difficult to replicate in an automated system, although the gap between wizard and computational capability has narrowed substantially during the course of this research (see Section 6.5).

The studies also took place in controlled laboratory settings within a single fabrication domain. Sewing involves particular characteristics: a loosely procedural task, material constraints, moderate safety concerns. Other domains, such as CNC machining, laser cutting, or large-scale industrial fabrication, involve different constraints and hazards that might change the dynamics of initiative. Real fabrication environments (homes, makerspaces, shared workshops) further introduce contextual complexity, social dynamics, and timescales absent from controlled settings.

6.5 Situated Against a Shifting AI Landscape

The years over which this research was conducted coincided with one of the most rapid transformations in AI capability and adoption since the field's founding. When this research began, the question of whether a fabrication machine could plausibly observe a maker, infer their state, and generate timely natural-language guidance was largely speculative. By the time this dissertation reaches its readers, those capabilities will be much closer to deployment.

Three developments are particularly relevant. First, multi-modal foundation models, including large vision-language models such as the GPT-4 family, Gemini, and Claude, can now describe scenes, identify common objects, and generate contextually appropriate language responses to complex visual inputs. The kinds of observations the wizard made in Chapter 5 ("the participant is looking around the side of the machine; they probably can't find the presser foot lever") are increasingly within the descriptive range of such models, particularly when paired with task-specific scaffolding.

Second, recent vision-language-action (VLA) models such as RT-2 [Zitkovich et al., 2023], OpenVLA [Kim et al., 2024], NVIDIA's GR00T [Bjorck et al., 2025], Physical Intelligence's π_0 [Black et al., 2024], and Google DeepMind's Gemini Robotics [Abeyruwan et al., 2025] extend foundation-model reasoning to embodied control, learning from large cross-embodiment datasets to map visual and linguistic inputs to physical action. While the present dissertation deliberately did not build a fully autonomous system, the broader research community is now actively building systems resembling the role of the wizard in Chapter 5: continuous observation, multi-modal interpretation, and generation of contextually appropriate output.

Third, the consumer fabrication landscape is under continuous development. Bambu Lab and similar manufacturers ship 3D printers with onboard cameras and neural-network-based first-layer inspection; sewing machines increasingly include sensors that adjust thread tension automatically; emerging consumer-facing assistants combine cameras and LLMs to comment on cooking, knitting, or assembly. As Chapter 3 noted, the next step along this trajectory would be fabrication machines that not only sense their own internal and task state, but also consider the user's state, and act on all three through language presented to users.

This shift sharpens the research questions of this dissertation rather than answering them. The bottleneck is no longer primarily whether a machine can describe what it sees or generate fluent language; it is whether such systems will know when to speak and how to calibrate to a maker's expertise. Horvitz [1999]'s principle that poorly timed initiative carries real costs is more relevant than ever: if a sufficiently capable model can produce plausible suggestions continuously, the design problem becomes one of restraint. The wizard in Chapter 5 was not impressive because of what she could say, but because of when she chose to say something, and when she chose to repeat in more detail. That judgment is the contribution this dissertation tries to characterize, and it is the part of the problem that recent VLA progress has not yet solved.

6.6 Future Directions

This research opens several directions for future work, organized below as the lines of inquiry introduced in Section 6.2, followed by two further directions on intervention modality and domain generalization.

6.6.1 Studies with Experts in Their Own Domains

The clearest next step is to replicate the Chapter 5 manipulation with experienced practitioners in the domains where they actually work. This is methodologically harder than working with novices: the wizard would need deeper domain expertise to make credible interventions, the task would need to be challenging at expert level, and recruitment would need to reach beyond university populations. A study comparing proactive and reactive assistance with, for example, advanced sewists working on garment construction, or with experienced 3D printing operators recovering from a failing print, could test whether the central finding (proactive assistance increases timeliness without increasing annoyance) holds, attenuates, or reverses with expertise.

6.6.2 Longitudinal Studies

A complementary direction is to follow the same users across the developmental trajectory P23 articulated. A study tracking participants from first use through progressive familiarity, across weeks or months in a domain like sewing, knitting, or woodworking, could clarify when proactive assistance crosses from welcome to unwelcome, what signals that transition, and whether users themselves can articulate the shift in advance. Longitudinal data could also speak to how users' mental models of machine capability develop over repeated interactions, and whether the novelty of a "talking machine" inflates initial positive responses.

6.6.3 Field Studies in Real Fabrication Contexts

The studies in this dissertation took place in controlled laboratory settings with equipment provided by the experimenter and tasks set by the protocol. Real fabrication happens in workshops, makerspaces, kitchens, and homes, with the user's own materials and on the user's own schedule, often with social context (a teacher, a peer, an audience) that the laboratory excludes. Whether the patterns observed here survive contact with these conditions is unknown. Goudswaard et al. [2024]'s observations on nonlinear activities during collaborative 3D printing offer one model for how field-grounded research can inform the design vision; an analogous project tracking proactive-assistance use in a public makerspace could help close the gap between lab studies and the messy contexts where fabrication actually happens. Such studies could also begin to address how machines should coordinate assistance among several users, and how social dynamics affect preferences for machine initiative, since both studies in this dissertation examined individual makers working alone.

6.6.4 Computational Detection of Intervention Opportunities

The wizard in Chapter 5 acted on behavioral patterns signaling confusion or hesitation, among other cues. The multi-view video datasets from Chapters 4 and 5 provide foundations for computational models that could detect these states automatically. The wizard's core capabilities of observing through cameras, interpreting behavioral cues, and generating contextually appropriate speech are increasingly within reach of current vision-language and vision-language-action models, as Section 6.5 discussed. Future work could annotate ground-truth hesitation moments and evaluate whether pose, gaze, and speech features can predict intervention opportunities with sufficient accuracy for real-time use.

The CoSew-4 and FabriCam-5 datasets were structured to support this work: synchronized multi-modal streams, labeled wizard interventions, and pose features chosen to surface the behavioral signals (stillness, gaze shift, hand-position drift) that pose-aware foundation models could be trained to recognize. A computational system could allow the questions raised in the preceding subsections to be asked at larger scales. Such systems would not need to outperform the wizard to be valuable; they would need only to be deployable across enough users, domains, and timescales that expert studies, longitudinal studies, and field studies become feasible.

6.6.5 Adaptive Initiative Systems

Chapter 5 primarily compared fixed initiative levels (reactive and proactive), aside from the escalation that took place before human intervention (see Section 5.3). Future systems might adjust dynamically based on detected expertise, task phase, or explicit user preferences. The developmental trajectory from instructive (for novices) to collaborative (for intermediate users) to responsive (for experts) that participants described suggests a model that adaptive systems could implement. Such adaptation is also the technical answer to the expert/novice gap of Section 6.2: a system that can scale its assistance to user expertise need not be optimized for one population at the expense of another.

6.6.6 Richer Intervention Modalities

Participants in Chapter 5 requested richer visual guidance for spatially complex tasks. The projected pointing mechanism helped, but some challenges, such as spatially orienting materials and understanding inside-out transformations, remained difficult to convey verbally. Future work could investigate how projected annotations, augmented reality overlays, or physical demonstrations complement verbal assistance. Recent advances in mixed-reality displays and in physically grounded video generation suggest that the modality space for fabrication guidance is widening.

6.6.7 Generalization Across Domains

The principles identified here should be tested across a broader range of fabrication contexts: 3D printing, laser cutting, CNC machining, and potentially domains beyond fabrication such as cooking, gardening, and musical performance. Each involves different constraints, hazards, and user populations that may require domain-specific adaptation.

6.7 Conclusion

This dissertation argues for a vision of human-machine collaboration in creative work, where machines serve as attentive, responsive partners rather than autonomous executors. We posit that achieving this vision requires understanding what makers value in the creative process, methods for capturing the signals that characterize human-machine fabrication, and empirical evidence about how users respond when machines take initiative.

The theoretical foundations that framed this research posed a tension. Sennett [2008] warned that automation risks reducing the relationship between maker and material, undermining the skill development and embodied engagement that make craft meaningful. Schön [2017] described skilled practice as a conversation with materials, one that requires the practitioner's active participation. Horvitz [1999] offered a framework for navigating this tension through mixed-initiative interaction, but emphasized that poorly timed initiative carries real costs. The question was whether these concerns would prove fatal to the idea of proactive fabrication machines, or whether a path existed between helpful support and unwanted automation.

The findings presented here suggest that such a path exists, at least for novice makers. When the machine anticipated confusion and offered guidance at the right moment, participants did not disengage from the material; when the machine confirmed their progress, they reported feeling less isolated and more motivated to persist through difficulty. The machine's confirmations ("You are doing it correctly") supported the reflective loop that Schön described: participants proposed moves, received feedback on whether the move was working, and adjusted. The machine became part of the reflective conversation.

At the same time, participants' anticipation that proactive help would become unwanted with experience is in line with Sennett's view. Skilled makers derive meaning from grappling with challenge, and a machine that continues to intervene after a maker has internalized the skill would undermine what makes the work satisfying. The design implication is not that machines should always help more, but that they should calibrate their involvement to the evolving relationship between maker and material by offering more intensive support during early encounters, and progressively stepping back as competence develops. Bridging from the novice evidence presented here to the expert vision that opened this dissertation requires the program of research outlined above: expert studies, longitudinal tracking, field deployments, and computational systems capable of supporting all three.

The challenge ahead is building the sensing and intelligence to achieve what the wizard demonstrated: recognizing hesitation, interpreting confusion, and deciding, in the moment, whether to speak or stay silent. The AI landscape that has developed during the course of this research has brought the components of such a system within reach. Vision-language models can describe what is happening in a fabrication scene; vision-language-action models can connect language to physical action; multi-modal foundation models can produce contextually appropriate speech. What none of these systems yet does well is decide *when* to act, which is the part of the wizard's role that this dissertation has tried to characterize. Translating the wizard's situated judgment into computational systems, without losing the sensitivity to human experience that made it effective, is the challenge for the next phase of this research.

This dissertation opened with three makers describing a tension in their work with machines. The crocheter valued the joy of learning a new skill with his hands, the potter wanted bumpers to prevent mistakes but took pride in making things herself, and the 3D printing expert wanted to reach into the machine and intervene mid-process. While the studies here do not answer the wishes of those experts, this dissertation lays groundwork toward addressing the gap, by illuminating the challenges and opportunities ahead. Our study participants echoed what the makers described as a desire not for more automation but for more attentive partnership: machines that step forward, that step back, and that know when a well-timed word of guidance is worth more than any amount of autonomous capability. The methods, empirical findings, datasets, and design principles in this dissertation contribute a foundation for mixed-initiative machines that help makers without presuming to automate their work.

BIBLIOGRAPHY

Saminda Abeyruwan, Joshua Ainslie, Jean-Baptiste Alayrac, Montserrat Gonzalez Arenas, Travis Armstrong, Ashwin Balakrishna, Robert Baruch, Maria Bauza, Michiel Blokzijl, et al. Gemini robotics: Bringing ai into the physical world. *arXiv preprint arXiv:2503.20020*, 2025.

Aegisub Contributors. Aegisub advanced subtitle editor, 2024. URL https://aegisub.org. Accessed: 2026-01-03.

Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, et al. Guidelines for human-ai interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems*, pages 1–13, 2019.

Mark Banks. Craft labour and creative industries. In *Creativity and cultural policy*, pages 75–91. Routledge, 2014.

Durand R Begault, Elizabeth M Wenzel, Richard Shrum, and Joel Miller. A virtual audio guidance and alert system for commercial aircraft operations. In *Proceedings of the Third International Conference on Auditory Display*, volume 117122, 1996.

Cindy L Bethel and Robin R Murphy. Survey of non-facial/non-verbal affective expressions for appearance-constrained robots. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 38(1):83–92, 2007.

Simone Bianco, Gianluigi Ciocca, Paolo Napoletano, and Raimondo Schettini. An interactive tool for manual, semi-automatic and automatic video annotation. *Computer Vision and Image Understanding*, 131:88–99, 2015.

Johan Bjorck, Fernando Castañeda, Nikita Cherniadev, Xingye Da, Runyu Ding, Linxi Fan, Yu Fang, Dieter Fox, Fengyuan Hu, Spencer Huang, et al. Gr00t n1: An open foundation model for generalist humanoid robots. *arXiv preprint arXiv:2503.14734*, 2025.

Kevin Black, Noah Brown, Danny Driess, Adnan Esmail, Michael Equi, Chelsea Finn, Niccolo Fusai, Lachy Groom, Karol Hausman, Brian Ichter, et al. *pi*_0: A vision-language-action flow model for general robot control. *arXiv preprint arXiv:2410.24164*, 2024.

Alan F Blackwell. Interacting with an inferred world: The challenge of machine learning for humane computer interaction. In *Proceedings of The Fifth Decennial Aarhus Conference on Critical Alternatives*, pages 169–180, 2015.

Susanne Bødker and Kaj Grønbæk. Users and designers in mutual activity: An analysis of cooperative activities in systems design. *Cognition and communication at work*, 1996.

Alexandra Bremers. A computer that sketches along with you. In *Creativity and Cognition*, C&C '22, New York, NY, USA, 2022a. doi: https://doi.org/10.1145/3527927.3533732. URL https://dl.acm.org/doi/abs/10.1145/3527927.3533732. event-place: Venice, Italy.

Alexandra Bremers. How can a robot help people draw? In *Companion Publication of the 2021 ACM Designing Interactive Systems Conference*, DIS '22 Companion, New York, NY, USA, 2022b. Association for Computing Machinery. doi: https://doi.org/10.1145/3532107.3532876. URL https://dl.acm.org/doi/abs/10.1145/3532107.3532876. event-place: Virtual Event.

Alexandra Bremers and Wendy Ju. Can machines tell what people want? bringing situated intelligence to generative ai. In *Proceedings of the Halfway to the Future Symposium*, pages 1–6, 2024a.

Alexandra Bremers and Wendy Ju. Designing interactions for mixed-initiative machines: Balancing automation and craftsmanship. In *International Conference on Human-Robot Interaction: Workshop on Worker-Robot Relations*, 2024b.

Alexandra Bremers, Alexandria Pabst, Maria Teresa Parreira, and Wendy Ju. Using Social Cues to Recognize Task Failures for HRI: A Review of Current Research and Future Directions, 2023a. URL https://arxiv.org/abs/2301.11972.

Alexandra Bremers, Maria Teresa Parreira, Xy Fang, Natalie Friedman, Adolfo Ramirez-Aristizabal, Alexandria Pabst, Mirjana Spasojevic, Mike Kuniavsky, and Wendy Ju. The Bystander Affect Detection (BAD) dataset for failure detection in HRI. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2023b. doi: https://doi.org/10.48550/arXiv.2303.04835.

Alexandra Bremers, Maria Teresa Parreira, and Wendy Ju. Understanding Bystander Facial Responses to Robot Task Failures with the BAD Dataset. In *HRI Workshop Position: The Imperfectly Relatable Robot*, 2023c.

Alexandra Bremers, Thijs Roumen, François Guimbretière, and Wendy Ju. FabriCam-5: Multiview Video Dataset of Novice Interactions with Tabletop Fabrication Machines. Harvard Dataverse, 2025. doi: 10.7910/DVN/X1LZKJ. URL https://doi.org/10.7910/DVN/X1LZKJ.

Alexandra WD Bremers, Natalie Friedman, Sam Lee, Tong Wu, Eric Laurier, Malte F Jung, Jorge Ortiz, and Wendy Ju. (social) trouble on the road: Understanding and addressing social discomfort in shared car trips. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces*, pages 1–13, 2024a.

Alexandra WD Bremers, Manaswi Saha, and Adolfo G Ramirez-Aristizabal. Situated conversational agents for task guidance: A preliminary user study. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces*, pages 1–7, 2024b.

Douglas AJ Brion and Sebastian W Pattinson. Generalisable 3d printing error detection and correction via multi-head neural networks. *Nature communications*, 13(1):4654, 2022.

Barry Brown, Fanjun Bu, Ilan Mandel, and Wendy Ju. Trash in motion: Emergent interactions with a robotic trashcan. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400703300. doi: 10.1145/3613904.3642610. URL https://doi.org/10.1145/3613904.3642610.

Fanjun Bu and Wendy Ju. Restory: Vlm-augmentation of social human-robot interaction datasets. In Oskar Palinko, Leon Bodenhagen, John-John Cabibihan, Kerstin Fischer, Selma Šabanović, Katie Winkle, Laxmidhar Behera, Shuzhi Sam Ge, Dimitrios Chrysostomou, Wanyue Jiang, and Hongsheng He, editors, *Social Robotics*, pages 457–466, Singapore, 2025. Springer Nature Singapore. ISBN 978-981-96-3525-2.

Fanjun Bu, Kerstin Fischer, and Wendy Ju. Making sense of robots in public spaces: A study of trash barrel robots. *J. Hum.-Robot Interact.*, 14(4), June 2025. doi: 10.1145/3731252. URL https://doi.org/10.1145/3731252.

Leah Buechley and Ruby Ta. 3D Printable Play-Dough: New Biodegradable

Materials and Creative Possibilities for Digital Fabrication. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–15, Hamburg Germany, April 2023. ACM. ISBN 978-1-4503-9421-5. doi: 10.1145/3544548.3580813. URL https://dl.acm.org/doi/10.1145/3544548.3580813.

Judee K. Burgoon, David B. Buller, Jerold L. Hale, and Mark A. de Turck. Relational messages associated with nonverbal behaviors. *Human Communication Research*, 10(3):351–378, 1984. doi: https://doi.org/10.1111/j.1468-2958.1984.tb00023.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-2958.1984.tb00023.x.

Bill Buxton. *Sketching user experiences: getting the design right and the right design*. Morgan kaufmann, 2010.

Nermin Caber, Jiaming Liang, Bashar I. Ahmad, Simon Godsill, Alexandra Bremers, Philip Thomas, David Oxtoby, and Lee Skrypchuk. Driver Profiling and Bayesian Workload Estimation for Adaptive In-Vehicle HMI. In *arXiv:2303.14720*, 2023. doi: https://doi.org/10.48550/arXiv.2303.14720.

Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Openpose: realtime multi-person 2d pose estimation using part affinity fields. *arXiv preprint arXiv:1812.08008*, 2018.

Timothy G. Clapp, Trevor J. Little, Theresa M. Thiel, and Dianna J. Vass. Sewing dynamics: Objective measurement of fabric/machine interaction. *International Journal of Clothing Science and Technology*, 4(2-3):45–53, 02 1992. ISSN 0955-6222. doi: 10.1108/eb002993. URL https://doi.org/10.1108/eb002993.

Matthew B Crawford. *Shop class as soulcraft: An inquiry into the value of work*. Penguin, 2009.

Nigel Cross. Designerly ways of knowing. *DESIGN STUDIES*, 3(4), 1982.

Mihaly Csikszentmihalyi. *Flow: The psychology of optimal experience*, volume 1990. Harper & Row New York, 1990.

Andrea Cuadra, Hansol Lee, Jason Cho, and Wendy Ju. Look at me when i talk to you: A video dataset to enable voice assistants to recognize errors. *arXiv preprint arXiv:2104.07153*, 2021.

Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg. Wizard of oz studies—why and how. *Knowledge-based systems*, 6(4):258–266, 1993.

Avital Dell'Ariccia, Alexandra Bremers, Johan Michalove, and Wendy Ju. How to make people think you're thinking if you're a drawing robot: Expressing emotions through the motions of writing. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction*, pages 1190–1191, 2022a.

Avital Dell'Ariccia, Alexandra WD Bremers, Wen-Ying Lee, and Wendy Ju. " ah! he wants to win!": Social responses to playing tic-tac-toe against a physical drawing robot. In *Sixteenth International Conference on Tangible, Embedded, and Embodied Interaction*, pages 1–6, 2022b.

Kees Dorst and Judith Dijkhuis. Comparing paradigms for describing design activity. *Design studies*, 16(2):261–274, 1995.

Dale Dougherty. The maker movement. *Innovations: Technology, governance, globalization*, 7(3):11–14, 2012.

Anca D Dragan, Kenton CT Lee, and Siddhartha S Srinivasa. Legibility and predictability of robot motion. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 301–308. IEEE, 2013.

Elsa Eiriksdottir and Richard Catrambone. Procedural instructions, principles, and examples: How to structure instructions for procedural tasks to enhance performance, learning, and transfer. *Human factors*, 53(6):749–770, 2011.

Michael Eraut. Schon shock: a case for refraining reflection-in-action? *Teachers and teaching*, 1(1):9–22, 1995.

K Anders Ericsson, Ralf T Krampe, and Clemens Tesch-Römer. The role of deliberate practice in the acquisition of expert performance. *Psychological review*, 100(3):363, 1993.

Evil Mad Scientist. *AxiDraw Python API Reference*. Evil Mad Scientist, 2025. URL https://axidraw.com/doc/py_api/. Accessed: 2025-10-12.

Evil Mad Scientist Wiki. Axidraw software installation. https://wiki.evilmadscientist.com/Axidraw_Software_Installation, 2024. URL https://wiki.evilmadscientist.com/Axidraw_Software_Installation. Last modified 29 February 2024, version v.3.9.5.

F. Flemisch, D. Abbink, M. Itoh, M-P. Pacaux-Lemoine, and G. Weßel. Shared control is the sharp end of cooperation: Towards a common framework of joint action, shared control and human machine cooperation. *IFAC-PapersOnLine*, 49(19):72–77, 2016. ISSN 2405-8963. doi: https://doi.org/10.1016/j.ifacol.2016.10.464. URL https://www.sciencedirect.com/science/article/pii/S2405896316320547. 13th IFAC Symposium on Analysis, Design, and Evaluation ofHuman-Machine Systems HMS 2016.

Natalie Friedman, Alexandra Bremers, Adelaide Nyanyo, Ian Clark, Yasmine Kotturi, Laura Dabbish, Wendy Ju, and Nikolas Martelaro. Understanding the challenges of maker entrepreneurship. volume 9, pages 1–29. ACM New York, NY, USA, 2025.

William Gaver. What should we expect from research through design? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 937–946, Austin Texas USA, May 2012. ACM. ISBN 978-1-4503-1015-4. doi: 10.1145/2207676.2208538. URL https://dl.acm.org/doi/10.1145/2207676.2208538.

Vlad Petre Glăveanu and Saadi Lahlou. Through the creator's eyes: Using the subjective camera to study craft creativity. *Creativity Research Journal*, 24(2-3): 152–162, 2012.

Gabriela Goldschmidt. *Linkography: unfolding the design process*. Mit Press, 2014.

Maas Goudswaard, Bruna Goveia Da Rocha, and Kristina Andersen. Entering the 3d printer: negotiations of imprecision in making. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*, DIS '24, page 1148–1161, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400705830. doi: 10.1145/3643834.3660758. URL https://doi.org/10.1145/3643834.3660758.

Itay Grinberg, Alexandra Bremers, Louisa Pancoast, and Wendy Ju. Implicit collaboration with a drawing machine through dance movements. In *Proceedings of the 8th ACM Symposium on Computational Fabrication*, SCF '23, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9798400703195. doi: 10.1145/3623263.3629150. URL https://doi.org/10.1145/3623263.3629150.

Renan Guarese, Emma Pretty, Aidan Renata, Deb Polson, and Fabio Zambetta. Exploring audio interfaces for vertical guidance in augmented reality via hand-based feedback. *IEEE Transactions on Visualization and Computer Graphics*, 30(5):2818–2828, 2024. doi: 10.1109/TVCG.2024.3372040.

Mahmoud Hassan, Ahmad Sadek, M Helmi Attia, and Vincent Thomson. Intelligent machining: real-time tool condition monitoring and intelligent adaptive control systems. *Journal of Machine Engineering*, 18(1):5–18, 2018.

David Hinwood, James Ireland, Elizabeth Ann Jochum, and Damith Herath. A proposed wizard of OZ architecture for a human-robot collaborative drawing task. In *International Conference on Social Robotics*, pages 35–44. Springer, 2018.

Guy Hoffman and Wendy Ju. Designing Robots with Movement in Mind. *J. Hum.-Robot Interact.*, 3(1):91–122, February 2014. doi: 10.5898/JHRI.3.1. Hoffman. URL https://doi.org/10.5898/JHRI.3.1.Hoffman. Publisher: Journal of Human-Robot Interaction Steering Committee.

Eric Horvitz. Principles of Mixed-Initiative User Interfaces. In *CHI '99: Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 159 – 166, 1999.

Eric J Horvitz, John S Breese, David Heckerman, David Hovel, and Koos Rommelse. The lumiere project: Bayesian user modeling for inferring the goals and needs of software users. *arXiv preprint arXiv:1301.7385*, 2013.

Interaction Design Foundation. What are semi-structured interviews? 2025. URL https://www.interaction-design.org/literature/topics/semi-structured-interviews. Accessed: 2025-04-30.

Chipp Jansen and Elizabeth Sklar. Exploring Co-creative Drawing Workflows. *Frontiers in Robotics and AI*, 8:577770, May 2021. ISSN 2296-9144. doi: 10.3389/frobt.2021.577770. URL https://www.frontiersin.org/articles/10.3389/frobt.2021.577770/full.

Brigitte Jordan and Austin Henderson. Interaction analysis: Foundations and practice. *The journal of the learning sciences*, 4(1):39–103, 1995.

Wendy Ju and Larry Leifer. The design of implicit interactions: Making interactive systems less obnoxious. *Design Issues*, 24(3):72–84, 2008. Publisher: MIT Press.

Jeeeun Kim, Haruki Takahashi, Homei Miyashita, Michelle Annett, and Tom Yeh. Machines as co-designers: A fiction on the future of human-fabrication machine interaction. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '17, page 790–805, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450346566. doi: 10.1145/3027063.3052763. URL https://doi.org/10.1145/3027063.3052763.

Moo Jin Kim, Karl Pertsch, Siddharth Karamcheti, Ted Xiao, Ashwin Balakrishna, Suraj Nair, Rafael Rafailov, Ethan Foster, Grace Lam, Pannag Sanketi, Quan Vuong, Thomas Kollar, Benjamin Burchfiel, Russ Tedrake, Dorsa Sadigh, Sergey Levine, Percy Liang, and Chelsea Finn. Openvla: An open-source vision-language-action model. *arXiv preprint arXiv:2406.09246*, 2024.

Dimosthenis Kontogiorgos, Minh Tran, Joakim Gustafson, and Mohammad Soleymani. A systematic cross-corpus analysis of human reactions to robot conversational failures. In *Proceedings of the 2021 International Conference on Multimodal Interaction*, pages 112–120, 2021.

Panagiotis N Koustoumpardis and Nikos A Aspragathos. Intelligent hierarchical robot control for sewing fabrics. *Robotics and Computer-Integrated Manufacturing*, 30(1):34–46, 2014.

Eric Laurier, Hayden Lorimer, Barry Brown, Owain Jones, Oskar Juhlin, Allyson Noble, Mark Perry, Daniele Pica, Philippe Sormani, Ignaz Strebel, et al. Driving and 'passengering': Notes on the ordinary organization of car travel. *Mobilities*, 3(1):1–23, 2008.

Eric Laurier, Barry Brown, and Lorimer Hayden. What it means to change lanes: actions, emotions and wayfinding in the family car. *Semiotica*, 2012(191):117–135, 2012.

Bryan Lawson. *How designers think*. Routledge, 2006.

Jamy Li, Andrea Cuadra, Brian Mok, Byron Reeves, Jofish Kaye, and Wendy Ju. Communicating dominance in a nonanthropomorphic robot using locomotion. *ACM Transactions on Human-Robot Interaction (THRI)*, 8(1):1–14, 2019.

Jinying Lin, Zhen Ma, Randy Gomez, Keisuke Nakamura, Bo He, and Guangliang Li. A review on interactive reinforcement learning from human social feedback. *IEEE Access*, 8:120757–120765, 2020a.

Yuyu Lin, Jiahao Guo, Yang Chen, Cheng Yao, and Fangtian Ying. It is your turn: Collaborative ideation with a co-creative robot through sketch. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–14, New York, NY, USA, 2020b. Association for Computing Machinery. ISBN 9781450367080. doi: 10.1145/3313831.3376258. URL https://doi.org/10.1145/3313831.3376258.

Wendy E Mackay, Raymonde Guindon, MM Mantel, Lucy Suchman, and Deborah G Tatar. Video: Data for studying human-computer interaction. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 133–137, 1988.

Nikolas Martelaro and Wendy Ju. Woz way: Enabling real-time remote interaction prototyping & observation in on-road vehicles. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*, pages 169–182. ACM, 2017.

Microsoft. Microsoft Platform for Situated Intelligence (Psi), 2024. URL https://github.com/microsoft/psi.

Brian Mok, Stephen Yang, David Sirkin, and Wendy Ju. Performing collaborative tasks with robotic drawers. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*, pages 309–309, 2015.

Stefanie Mueller, Pedro Lopes, Konstantin Kaefer, Bastian Kruck, and Patrick Baudisch. constructable: interactive construction of functional mechanical devices. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pages 3107–3110. 2013.

Stefanie Mueller, Sangha Im, Serafima Gurevich, Alexander Teibrich, Lisa Pfisterer, François Guimbretière, and Patrick Baudisch. Wireprint: 3d printed previews for fast prototyping. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14, page 273–280, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450330695. doi: 10.1145/2642918.2647359. URL https://doi.org/10.1145/2642918.2647359.

Clifford Nass, Jonathan Steuer, and Ellen R Tauber. Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 72–78, 1994.

Jennifer Ockerman and Amy Pritchett. A review and reappraisal of task guidance: Aiding workers in procedure following. *International Journal of Cognitive Ergonomics*, 4(3):191–212, 2000.

Dan R Olsen and Michael A Goodrich. Metrics for evaluating human-robot interactions. In *Proceedings of PERMIS*, volume 2003, page 4, 2003.

OpenAI. openai/whisper: Robust speech recognition via large-scale weak supervision. https://github.com/openai/whisper, 2022. URL https://github.com/openai/whisper. Accessed: 2025-10-16.

Nadya Peek, James Coleman, Ilan Moyer, and Neil Gershenfeld. Cardboard machine kit: Modules for the rapid prototyping of rapid prototyping machines. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 3657–3668, 2017.

Hannah Pelikan. Transcribing human–robot interaction: Methodological implications of participating machines. In *Ethnomethodological Conversation Analysis in Motion*, pages 42–62. Routledge, 2023.

Hannah R. M. Pelikan and Malte F. Jung. Designing robot sound-in-interaction: The case of autonomous public transport shuttle buses. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '23, page 172–182, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9781450399647. doi: 10.1145/3568162.3576979. URL https://doi.org/10.1145/3568162.3576979.

Hannah R. M. Pelikan, Mathias Broth, and Leelo Keevallik. ”are you sad, cozmo?”: How humans make sense of a home robot’s emotion displays. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, HRI ’20, page 461–470, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450367462. doi: 10.1145/3319502.3374814. URL https://doi.org/10.1145/3319502.3374814.

Huaishu Peng, Jimmy Briggs, Cheng-Yao Wang, Kevin Guo, Joseph Kider, Stefanie Mueller, Patrick Baudisch, and François Guimbretière. Roma: Interactive fabrication with augmented reality and a robotic 3d printer. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI ’18, page 1–12, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450356206. doi: 10.1145/3173574.3174153. URL https://doi.org/10.1145/3173574.3174153.

Martin Porcheron, Joel E Fischer, and Stuart Reeves. Pulling back the curtain on the wizards of oz. *Proceedings of the ACM on Human-Computer Interaction*, 4 (CSCW3):1–22, 2021.

Robin Jephthah Rajarathinam, Christian Palaguachi, and Jina Kang. Enhancing multimodal learning analytics: A comparative study of facial features capture using traditional vs 360-degree cameras in collaborative learning. *NSF Public Access Repository*, 2024.

Laurel D Riek. Wizard of oz studies in hri: a systematic review and new reporting guidelines. *Journal of Human-Robot Interaction*, 1(1):119–136, 2012.

Mose Sakashita, E. Andy Ricci, Jatin Arora, and François Guimbretière. Remotecode: Robotic embodiment for enhancing peripheral awareness in remote collaboration tasks. *Proc. ACM Hum.-Comput. Interact.*, 6(CSCW1), apr 2022. doi: 10.1145/3512910. URL https://doi.org/10.1145/3512910.

Giulio Sandini and Alessandra Sciutti. Humane robots—from robots with a humanoid body to robots with an anthropomorphic mind. *J. Hum.-Robot Interact.*, 7(1), may 2018. doi: 10.1145/3208954. URL https://doi.org/10.1145/3208954.

Peter Schaldenbrand, Gaurav Parmar, Jun-Yan Zhu, James McCann, and Jean Oh. Cofrida: Self-supervised fine-tuning for human-robot co-painting. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024.

Donald A Schön. *The reflective practitioner: How professionals think in action*. Routledge, 2017.

Evil Mad Scientist. Axidraw v3. https://shop.evilmadscientist.com/, 2022.

Rob Semmens, Nikolas Martelaro, Pushyami Kaveti, Simon Stent, and Wendy Ju. Is now a good time?: An empirical study of vehicle-driver communication timing. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, page 637, New York, NY, USA, 2019. ACM, ACM.

Richard Sennett. *The craftsman*. Yale University Press, 2008.

Thomas B Sheridan. *Telerobotics, automation, and human supervisory control*. MIT press, 1992.

David Sirkin, Brian Mok, Stephen Yang, and Wendy Ju. Mechanical ottoman: how robotic furniture offers and withdraws support. In *Proceedings of the Tenth*

Annual ACM/IEEE International Conference on Human-Robot Interaction, pages 11–18, 2015.

KG Srinivasa, Sriram Anupindi, R Sharath, and S Krishna Chaitanya. Analysis of facial expressiveness captured in reaction to videos. In *2017 IEEE 7th International Advance Computing Conference (IACC)*, pages 664–670, New York, NY, USA, 2017. IEEE, IEEE.

B Srinivasa Prasad, D Siva Prasad, A Sandeep, and G Veeraiah. Condition monitoring of cnc machining using adaptive control. *International Journal of Automation and Computing*, 10(3):202–209, 2013.

Blair Subbaraman, Nathaneal Bursch, and Nadya Peek. It's not the shape, it's the settings: Tools for exploring, documenting, and sharing physical fabrication parameters in 3d printing. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400713941. doi: 10.1145/3706598.3713354. URL https://doi.org/10.1145/3706598.3713354.

Lucy A Suchman. *Plans and situated actions: The problem of human-machine communication*. Cambridge university press, 1987.

Lucy A Suchman and Randall H Trigg. Understanding practice: Video as a medium for reflection and design. In *Design at work*, pages 65–89. CRC Press, 2020.

Anna Syberfeldt, Oscar Danielsson, Magnus Holm, and Lihui Wang. Visual assembling guidance using augmented reality. *Procedia Manufacturing*, 1:98–109, 2015.

Leila Takayama, Doug Dooley, and Wendy Ju. Expressing thought: improving robot readability with animation principles. In *Proceedings of the 6th international conference on Human-robot interaction*, pages 69–76, Lausanne Switzerland, March 2011. ACM. ISBN 978-1-4503-0561-7. doi: 10.1145/1957656.1957674. URL https://dl.acm.org/doi/10.1145/1957656.1957674.

Theresa Jean Tanenbaum, Amanda M Williams, Audrey Desjardins, and Karen Tanenbaum. Democratizing technology: pleasure, utility and expressiveness in diy and maker practice. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 2603–2612, 2013.

Tock Custom. How to sew pillows! easy beginner sewing tutorial for beginners. https://www.youtube.com/watch?v=lp5kUr7U9fE, September 2022. YouTube video, published September 8, 2022.

Stefan Trausan-Matu and James D Slotta. Artifact analysis. In *International handbook of computer-supported collaborative learning*, pages 551–567. Springer, 2021.

Yan Wang, Wei Song, Wei Tao, Antonio Liotta, Dawei Yang, Xinlei Li, Shuyong Gao, Yixuan Sun, Weifeng Ge, Wei Zhang, et al. A systematic review on affective computing: Emotion models, databases, and recent advances. *Information Fusion*, 2022.

Karl DD Willis, Cheng Xu, Kuan-Ju Wu, Golan Levin, and Mark D Gross. Interactive fabrication: new interfaces for digital fabrication. In *Proceedings of the fifth international conference on Tangible, embedded, and embodied interaction*, pages 69–72, 2010.

Margaret Wilson. Six views of embodied cognition. *Psychonomic bulletin & review*, 9(4):625–636, 2002.

Fei Wu, Jerald Thomas, Shreyas Chinnola, and Evan Suma Rosenberg. Comparison of audio and visual cues to support remote guidance in immersive environments. In *ICAT-EGVE*, pages 121–130, 2020.

Yi Xiong, Yunlong Tang, Samyeon Kim, and David W Rosen. Human-machine collaborative additive manufacturing. *Journal of Manufacturing Systems*, 66: 82–91, 2023.

JD Zamfirescu-Pereira, David Sirkin, David Goedicke, Ray LC, Natalie Friedman, Ilan Mandel, Nikolas Martelaro, and Wendy Ju. Fake it to make it: Exploratory prototyping in hri. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, pages 19–28, 2021.

John Zimmerman, Jodi Forlizzi, and Shelley Evenson. Research through design as a method for interaction design research in hci. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 493–502, 2007.

Brianna Zitkovich, Tianhe Yu, Sichun Xu, Peng Xu, Ted Xiao, Fei Xia, Jialin Wu, Paul Wohlhart, Stefan Welker, Ayzaan Wahid, Quan Vuong, Vincent Vanhoucke, Huong Tran, Radu Soricut, Anikait Singh, Jaspiar Singh, Pierre Sermanet, Pannag R. Sanketi, Grecia Salazar, Michael S. Ryoo, Krista Reymann, Kanishka Rao, Karl Pertsch, Igor Mordatch, Henryk Michalewski, Yao Lu, Sergey Levine, Lisa Lee, Tsang-Wei Edward Lee, Isabel Leal, Yuheng Kuang, Dmitry Kalashnikov, Ryan Julian, Nikhil J. Joshi, Alex Irpan, Brian Ichter, Jasmine Hsu, Alexander Herzog, Karol Hausman, Keerthana Gopalakrishnan, Chuyuan Fu, Pete Florence, Chelsea Finn, Kumar Avinava Dubey, Danny

Driess, Tianli Ding, Krzysztof Marcin Choromanski, Xi Chen, Yevgen Chebotar, Justice Carbajal, Noah Brown, Anthony Brohan, Montserrat Gonzalez Arenas, and Kehang Han. Rt-2: Vision-language-action models transfer web knowledge to robotic control. In Jie Tan, Marc Toussaint, and Kourosh Darvish, editors, *Proceedings of The 7th Conference on Robot Learning*, volume 229 of *Proceedings of Machine Learning Research*, pages 2165–2183. PMLR, 06–09 Nov 2023. URL https://proceedings.mlr.press/v229/zitkovich23a.html.

Victor W Zue and James R Glass. Conversational interfaces: Advances and challenges. *Proceedings of the IEEE*, 88(8):1166–1180, 2000.

APPENDIX A

MACHINE OPERATOR INTERVIEW QUESTIONS (CHAPTER 3)

We are going to ask questions about your press brake usage and how you interact with the machine to achieve those tasks.

1. *Can you tell us about yourself and how you started working in metal fabrication?*
   (a) *When did you decide to work in metal fabrication?*
   (b) *Why did you decide to work in metal fabrication?*
   (c) *What was your experience before working here?*
   (d) *How do the daily tasks in metal fabrication compare?*
2. *Walk us through a normal day at Vernier Metal Fabrication in terms of your main tasks, actions, and interactions.*
   (a) *Context:*
      i. *What is your schedule like?*
      ii. *Do you always work on-site?*
      iii. *How collaborative is the work environment?*
      iv. *What are the conditions usually like when you perform a bending task (environmental noise, crowdedness of workplace, space required)?*
   (b) *Task Process:*
      i. *Can you walk us step-by-step through the actions you perform when you are completing a metal bending task?*
      ii. *How are you alerted of a new bending job?*
      iii. *How do you prepare for the bending task? Do you need to prepare the environment or arrange objects, materials or machines differently?*

iv. *What are your three most-favorite and three least-favorite aspects or parts of performing the task?*

(c) *Tools:*

i. *Do you use any tools? How do you use them?*

ii. *Do you use any interfaces on the machine? How do you use them?*

iii. *Is there anything you wish would be different about the tools and machines that you are using to perform metal bending?*

3. *Can you think back to the last time you needed to bring someone new on to use the press brake machine?*

(a) *What was the background of metal working experience of the person who you were instructing?*

(b) *What did you have to teach them with respect to bending?*

(c) *How did you approach the instruction?*

(d) *Did you use any external tools (such as instruction videos) as part of the instruction process?*

(e) *Were there any parts of the process that the new user was particularly struggling with?*

(f) *Were there any parts of the process that the new user expressed being nervous or confused about?*

(g) *Are there any deciding factors that conclude the onboarding process?*

(h) *Are there any deciding factors that determine that the new worker can be assigned tasks to complete independently?*

(i) *Is there anything that comes to mind that you think could help improve the onboarding process for new workers?*

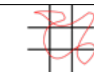
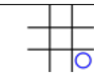
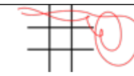
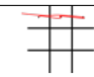
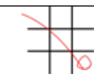
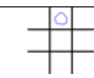
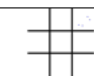
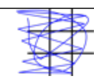
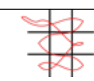
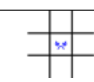
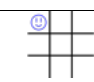
4. *Was there anything about using the press brake machine that you maybe forgot to mention or that we failed to ask that would help us understand the nature of your work?*

5. *Were there other actions we did not cover? [repeat set of questions for them]*

APPENDIX B

EXPRESSIVE AXIDRAW MOVEMENTS IN TIC-TAC-TOE (CHAPTER 3)

Table B.1: Expressive AxiDraw movements for Tic-Tac-Toe. (Previously published in Dell'Ariccia et al. [2022a]).

Movement	**Characteristics**	**SVG**	**Pen**
Slowly wandering around grid before writing.	Indirect movements and low speed.		Up
Confident, fast, no wandering.	Direct movement and high speed, urgent, pen down.		Down
Quickly wandering during opponent's turn to distract.	Indirect movements and high speed.		Up
Slowly moving around a few places.	Semi-direct movements and slow speed.		Up
Slow, roaming around same place repetitively.	Indirect movement and low speed.		Up
Not confident, shivering while writing.	Slow movement, semi-direct, pen down.		Down
"Cheating", attempt to write over opponent's move.	Indirect, high velocity, high acceleration, pen down.		Up
Mad, scribbling all over the place, chaotic motion.	Indirect, high speed, high acceleration, pen down.		Down
Happy, moves around to celebrate.	Indirect, high speed.		Up
Furious, draws angry emoji on grid.	Direct, high speed, high acceleration, pen down.		Down
Cheerful, cocky, draws happy emoji in winning place.	Direct, high speed, pen down.		Down

APPENDIX C

EXPRESSIVE AXIDRAW MOVEMENTS FROM DANCE (CHAPTER 3)

Table C.1: Welcoming stage: potential interactions before collaborative drawing. (Previously published in Grinberg et al. [2023].)

Interaction (robot and user)	**Implementation (robot)**
The robot calls you to come here. The robot wants you to notice it.	Make noise, wave.
The robot makes eye contact. The robot points at you.	Point pen at person and follow their gaze.
The robot invites you to sit down.	Point pen at person, then at chair, and back at person.
If the wrong person sits down, the robot tries to communicate: ”No, not him – you!”	Shake the pen and then point at the right person.

Table C.2: Collaboration stage: interactions during collaborative drawing. (Previously published in Grinberg et al. [2023].)

Interaction (robot and user)	**Implementation (robot)**
The robot asks what kind of card you want to make.	Point at cards and back at person.
The robot asks for paper.	Point at paper, participant, and back to paper.
The robot might start drawing.	Alternate between plotting and emotive movements.
The robot wants you to draw.	Point at person, at the paper, then again at the person.
You draw.	Wait in the home corner and make small observing movements.
The robot wants to see what you have drawn.	Hover over the paper before proceeding to add plotted elements.
The robot thinks the design is done.	Plot the ”Axi” signature, bow and wait for the participant to take the paper.
The robot thinks you should take the drawing away.	Point at paper, point at person, and bow again.

APPENDIX D

INTERVIEW QUESTIONS (CHAPTER 4)

1. What are your first thoughts on the experiment we just completed?
2. What was your prior level of experience with each of these tasks?
    - Sewing
    - Watercolor
    - Ceramics
3. Did you feel like you could do the tasks as you naturally would, despite being recorded?
4. Were there any specifically enjoyable moments doing these tasks?
5. Were there any specifically difficult moments, where the task was challenging?
6. Were there any confusing moments?
7. Were there any frustrating moments?
8. How would you describe your own performance on these tasks?
9. Are there any changes to the setup that you recommend that would make the tasks more natural for you? For instance, think of the fact that you were seated?
10. Are there any parts, devices, or tools that you would like to add to these tasks? For instance, other materials for sewing, etc.?
11. Do you have any other comments or suggestions for us?

APPENDIX E

QUALITATIVE CODING (CHAPTER 4)

Table E.1: Counts of codes after initial exploratory analysis of the FabriCam-5 dataset.

Code	Count
Experimenter Helps	27
Experimenter Instructs	30
Experimenter Sets Up	13
Participant Asks for Help	39
Participant Commands	20
Participant Curses	12
Participant Finds Solution	5
Participant Finishes Cleaning	7
Participant Gestures	1
Participant Headshake	1
Participant Helps Experimenter	1
Participant Hesitates	29
Participant Says "Hmm"	12
Participant Looks Closely	52
Participant Maneuvers / Adjusts	7
Participant Makes a Mistake	37
Participant Nods	2
Participant Says "Done"	2
Participant Signs / Signature	1
Participant Smiles	34
Participant Smirks	11
Participant Thanks	3
Participant Unexpected Reaction / Surprise	13
Total	**359**

APPENDIX F

POST-TASK QUESTIONNAIRE (CHAPTER 5)

Questionnaire after Task 1

Statement	**Strongly Disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**
The machine help was **clear**.	☐	☐	☐	☐	☐
The machine help was **helpful**.	☐	☐	☐	☐	☐
The machine help was **annoying**.	☐	☐	☐	☐	☐
The machine help was **trustworthy**.	☐	☐	☐	☐	☐
The machine help was **timely**.	☐	☐	☐	☐	☐

Comments: ____

____

Questionnaire after Task 2

Statement	**Strongly Disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**
The machine help was **clear**.	☐	☐	☐	☐	☐
The machine help was **helpful**.	☐	☐	☐	☐	☐
The machine help was **annoying**.	☐	☐	☐	☐	☐
The machine help was **trustworthy**.	☐	☐	☐	☐	☐
The machine help was **timely**.	☐	☐	☐	☐	☐

Comments: ____

____

APPENDIX G

INTERVIEW QUESTIONS (CHAPTER 5)

Intro Script

Thank you for participating in this study. We will now ask you a few questions about your experience you just had interacting with the machine, and also about your experiences with sewing in general. Your responses will be completely anonymized. Do you give us consent to video record this interview? (if yes, I will start recording now)

Interview Questions

- What are your first thoughts on the activities just completed?
- What are your thoughts on your interactions with the machine?
- What was your prior level of experience with sewing?
- How would you describe your own performance on the task?
- Were there any specifically enjoyable moments doing these tasks?
- Were there any specifically difficult or frustrating moments?
- Are there any changes that you recommend that would make the task better for you?
- Do you wish for different ways of interacting with tools and machines, than what is currently available?
- What are some things you wish a tool or machine could help you with?
- Would you be interested in using tools that take more of an initiative, and become a collaborator?
- Do you have any other comments or suggestions for us?

Closing

Thank you so much for completing the interview with us. Your input is really valuable to our research! Please feel free to contact us if any questions or concerns come up after the interview.

APPENDIX H

SETUP PHOTOS (CHAPTER 5)

Figure H.1: Left: The wizard sits behind the white screen and has access to a computer streaming the two webcams, an iPad streaming the 360 degree footage, and a phone streaming the eyeglass camera (not pictured). Right: above the participant's desk a projector is mounted. To the left, two example plushies are provided as a reference.