Homaeian, L., Wallace, J.R., Scott, S.D (2021). Joint Action Storyboards: A Framework for Visualizing Communication Grounding Costs. Proceedings of the ACM on HCI, (in press, accepted Jan 2021)

Joint Action Storyboards: A Framework for Visualizing Communication Grounding Costs

Leila Homaeian[1], leila.homaeian@uwaterloo.ca

James R. Wallace[2], jame.wallace@uwaterloo.ca

Stacey D. Scott[1,3], stacey.scott@uoguelph.ca

[1] Systems Design Engineering, University of Waterloo, Waterloo, Ontario, Canada

[2] School of Public Health and Health Systems, University of Waterloo, Waterloo, Ontario, Canada

[3] School of Computer Science, University of Guelph, Guelph, Ontario, Canada

ABSTRACT

Building and maintaining common ground is vital for effective collaboration in CSCW. Moreover, subtle changes in a CSCW user interface can significantly impact grounding and collaborative processes. Yet, researchers and technology designers lack tools to understand how specific user interface designs may hinder or facilitate communication grounding. In this work, we leverage the well-established theory of communication grounding to develop a visual framework, called Joint Action Storyboards (JASs), to analyze and articulate how interaction minutiae of a CSCW environment impact the costs of communication grounding. JASs can depict an integrated view of mental actions of collaborators, their physical interactions with each other and the CSCW environment, and the corresponding grounding costs incurred. We present the development of JASs and discuss its various benefits for HCI and CSCW research. Through a series of case studies, we demonstrate how JASs provide an analysis tool for researchers and technology designers and serve as a tool to articulate the impact of interaction minutiae on communication grounding.

CCS CONCEPTS

• Human-centered computing • Collaborative and social computing• Collaborative and social computing design and evaluation methods

KEYWORDS

CSCW; groupware; CSCW frameworks, communication grounding; grounding costs; design evaluation methods, storyboards

1 Introduction

Previous research in Human-Computer Interaction (HCI) asserts the critical role of building and maintaining common ground, or shared beliefs, assumptions, and understanding for effective collaboration [5]. In their seminal work, Clark and Brennan [10] introduced communication grounding as the dynamic and interactive process of establishing and maintaining common ground conversational parties. For example, during collaboration, people's common ground includes what they both know about each other, the environment, and the task. Each collaborator then formulates utterances or executes actions based on what they expect the other parties to already know [10]. The communication grounding framework conceptualizes characteristics or affordances of communication media that impact language use and therefore communication grounding, e.g., visibility and audibility. When a medium lacks one affordance, people use alternative, and often more effortful, approaches for grounding [18,19]. Clark and Brennan [10] also defined a set of grounding costs that can impact the effort required to establish common ground, depending on the affordances of a given medium. The communication grounding framework has been indispensable to the HCI community in understanding and predicting the costs and benefits of communication media [4,17,35], and has been extended to include more recent advances in computing like tangibility and mobility [27] and multitasking ability and awareness [16].

However, as communication and collaboration technologies become more complex and vary in form and function, including personal and shared devices like smartphones, tablets, and large wall and tabletop displays, the application of Clark and Brennan's original framework becomes less clear [35]. Moreover, technological advancements might hinder communication grounding in synchronous collaborative settings even though people can readily use communication channels such as speech [10,11,23]. Affordances lacking on one device may be compensated for one another. Interactions between multiple devices and/or people may involve a myriad of tradeoffs between their affordances and consequential grounding costs. Meanwhile, even small subtle changes to a CSCW user interface can significantly impact collaboration [21]. So, it is often unclear how a specific interface or interaction design may impact communication grounding and ultimately collaborative processes [11,12,23,26]. Thus, researchers lack tools to examine the specific impact of the interface design of a CSCW system on the grounding process to help uncover the positive and negative effects of an interface design on collaboration.

To address this need, we developed a visual framework, called Joint Action Storyboards (JASs), that adapts the concept of joint actions from Clark's original framework [9] to break down momentary, collective actions between participants in synchronous collaboration, whether remote or collocated. JASs provide a tool to understand and communicate how subtle interface designs can hinder or support communication grounding. JASs can also help identify potential opportunities to improve the design of the interface to better support collaboration. Clark's notion of joint actions focuses on social processes (primarily verbal exchanges) and mental actions of conversational parties [9]. In the context of CSCW, we expand this notion by including physical interactions with the environment. Therefore, by analyzing collaborators' joint actions, researchers and technology designers can see mental and social processes as well as physical interactions that must take place to update the group's common ground [9,35]. The mental actions underlie physical interactions (with collaborators or technology), which serve to mediate communication and coordination [9,35]. JASs are designed to focus on momentary actions, which we call *instances of interest*, as the unit of analysis. While brief in time, these instances of interest are typically joint actions that occur frequently during a collaboration session and are critical to the progression of a collaborative task. Thus, they have the potential to significantly impact the grounding process. The visual nature of JASs allows one to depict the context, for example, user interface elements and spatial positioning of collaborators, in which an instance of interest takes place for the analyst to take into account [32]. Moreover, the visualizations help generate empathy [32].

Our work is a continuation of efforts by other CSCW and HCI researchers to provide frameworks and analytic tools that can help us understand, describe, and analyze CSCW around rapidly-evolving technological systems, for instance, Hierarchical Task Analysis [40], Cognitive Task Analysis [30,43], Hybrid Cognitive Task Analysis [37], and Groupware Task Analysis [45]. Compared with JASs, such frameworks and tools either focus on modeling expert user behaviour or have a coarser analysis granularity focused on high-level concepts related to task or collaborative workflow. Thus, they are not suitable for uncovering how specific design features can impact low-level collaboration processes and interaction minutiae. Additionally, although effective CSCW relies on communication grounding [5,24], existing frameworks fall short on helping researchers and technology designers understand if a given design supports or impedes communication grounding [11,12]. The JASs framework provides a fine-grained analysis tool to understand and articulate the impact of a specific technique on communication grounding at the user interface level.

JASs may be used as a discount evaluation method, for instance, to assess how well a technology supports or hinders communication grounding prior to running a user study and by considering potential technology use cases. JASs may also be used alongside other methods, for example, Clark and Brennan's affordances framework [10], questionnaires, conversational analysis, thematic analysis, and observations, to study the impact of technology on communication grounding to provide a holistic view of the system. Those existing methods usually capture data that are observable in the physical world, making them appropriate for identifying certain instances where technology potentially impacts communication grounding. However, communication grounding includes mental actions that take place inside collaborators' minds, making them difficult to identify and analyze using the above methods. The concept of joint actions adapted in JASs captures such internal mental actions, allowing for a fine-grained analysis of grounding costs. JASs are also applicable to the design stage of a project as a discount method. In this case, one may identify instances of interest by creating hypothetical collaborative scenarios. The framework can then be used as a communication tool between technology designers and evaluators of the CSCW system.

We present the development of JASs through a case study in which we analyze grounding costs incurred by two existing cross-device interaction techniques, called TOUCH and TILT, designed by Homaeian et al. [23] to support co-located collaborative sensemaking around geospatial data. TOUCH provided a touch-based approach for selecting data on a shared, large tabletop to view on a connected personal display, while TILT involved remotely selecting the data using on-board motion sensors on the personal display. Homaeian et al.'s study of these techniques found that TOUCH afforded effective tightly coupled work, whereas TILT better afforded independent data exploration (loosely coupled work). However, their findings, derived from thematic analysis of their collected video data, did not articulate precisely *how* or *why* the respective interface techniques impacted low-level collaboration processes, such as communication grounding. As technology designers, we sought a more fined-grained analysis method that would help us identify and articulate, to ourselves and others, how specific design features (i.e. user interface and interaction designs) of a CSCW system impacted collaboration minutiae, to help us improve our designs. This goal inspired us to explore grounding costs imposed by the two techniques, and to understand the subtle impact of the two design approaches on communication grounding. We developed JASs while trying to probe these differences. We then apply JASs to a series of collaborative case studies to demonstrate the framework's key benefits for HCI and CSCW research.

In summary, we contribute JASs, a visual framework to analyze grounding costs in synchronous collaborative technologies (remote or co-located) that provides HCI researchers and technology designers with an integrated tool to

1. depict collaborators' mental actions, interactions with CSCW technologies, and the physicality of the system,
2. articulate the impact of specific user interface elements of CSCW technologies on grounding costs,
3. evaluate the impact of specific user interface elements on communication grounding as an analytical tool, and
4. analyze interaction minutiae to reveal potential design opportunities.

2 Related Work

The ability of team members to establish common ground is crucial to effective collaboration in synchronous CSCW systems [5]. This is especially true during mixed-focus work, when people come together after periods of individual work to share and merge their findings as a group [24]. In these settings, grounding spans multiple collaborators and often multiple devices, and includes mental actions and physical interactions with colleagues and the CSCW system., and is, therefore, a complicated and difficult process to analyze [11,12]. It is important to take into account interaction minutiae and breakdowns in the analysis to understand how subtle interface design choices support or impede grounding [10,21]. Yet, given the ubiquity of multi-user and multi-device environments, there is a pressing need for CSCW and HCI researchers to develop new tools that can aid the research and design of grounding in these settings [11,12,17]. Such tools can facilitate and fill gaps in research on more complex multi-display environments by the CSCW community [46].

Existing tools tend to fall short of providing this aid by including interactions among people only, modeling error-free behaviour, or focusing on a whole system rather than specific interface design elements. For example, CSCW and HCI researchers often use conversational coding and questionnaires (e.g., [12,15,20,25]) to study the impact of specific technological settings on communication grounding. These research tools are useful to understand communication structure and participants' perception of the impact of technology on communication grounding. However, they do not capture physical interactions with technology or the underlying mental actions, which are fundamental to the grounding process and are impacted by user interface features [10,35].

Conversely, user- and task-level analysis methods provide a means of understanding interactions with computer interfaces but are designed to model error-free interactions with computer systems. Predictive models like GOMS [13,40], The Keystroke Level Model [7,40], Touch Level Model [41], and task analysis methods such as Hierarchical Task Analysis [40] offer high precision in understanding these physical interactions (and in some cases cognitive processes), but are largely designed to analyze single user contexts and fail to account for interaction breakdowns. Therefore, such methods are not effective for studying user interface designs of CSCW technology that get in the way of the grounding process. Cognitive Task Analysis [30,43], Hybrid Cognitive Task Analysis [37], and Groupware Task Analysis [45] provide tools to understand collaborative workflows and consider cognitive overload during interactions with an interface. However, they have a broader unit of analysis compared to that of JASs. They focus on a whole system or a general mission and therefore are not suitable for studying interaction minutiae. Discount evaluation methods, such as Heuristic Evaluation and Walkthroughs [40] may be adapted to inspect usability problems in CSCW systems, for example by using the mechanics of collaboration [39] to model a collaborative task. However, those methods also have a broader unit of analysis and thus may miss to identify opportunities for a design improvement to alleviate grounding costs. Moreover, JASs are not only an analytical tool but also a communication tool to articulate the impact of user interface elements on collaborative processes.

In this work, we develop an analytical framework that helps researchers and designers examine mental actions and interactions with CSCW technology to analyze how certain interface design elements facilitate or impede communication grounding. We decided to focus on understanding grounding, and the grounding costs incurred with specific interfaces, first in a co-located CSCW setting. We chose this setting as a starting point for our investigations because grounding costs are generally low in face-to-face settings [10]; yet, co-located CSCW technologies such as multi-display environments can introduce new complexities that can hinder communication and collaboration processes [42]. For example, sharing information across displays in a multi-display system to facilitate a joint discussion can potentially interfere with the flow of conversation, and thus communication grounding, if a cross-device interaction technique is overly cumbersome.

In developing this analytical framework, we bridged three distinct areas of CSCW research: communication grounding theory, analytical frameworks, and theories for understanding collaborative processes.

2.1 Communication grounding theory

The theory of communication grounding [10] has been widely used by HCI researchers to predict and understand how a given technology may support or impede communication [4,35]. It describes the low-level processes that humans use during communication to establish and maintain common ground throughout a conversation. It this work, Clark and Brennan [10] describe a number of costs that can be incurred by the speaker and/or the listener during a conversation. These costs are context dependent and rely heavily on the setting in which the conversation occurs. For instance, in a face-to-face environment where conversational partners with no visual or auditory impairments have access to rich verbal and non-verbal communication channels it is easy to convey intended meaning and notice misunderstandings. Thus, for many types of conversations, face-to-face is considered the ideal grounding environment. On the other hand, when holding a conversation over the phone, speakers and listeners have no access to non-verbal communication, for instance, to convey a wry smile with a sarcastic comment or interpret the reaction that a message evoked.

These grounding costs are impacted by different communication affordances that distinct communication contexts have, including, for instance, whether or not parties can see each other (visibility), hear each other (audibility), talk at the same time (simultaneity), review what was said (reviewability), and so on [10]. These affordances can be thought of as 'resources for grounding' [35]. That is, the more affordances a given communication context provides, the better it will generally be at supporting grounding.

However, as CSCW technologies rapidly evolve, using this original framework to predict or explain problems for a given system becomes less clear [35]. For example, consider a multi-device co-located environment that has personal and shared devices. This environment has visibility, but people's attention can be focused on personal devices and therefore others' actions may not be immediately visible to them. Moreover, although Clark's theory of communication grounding is valuable to guide the design of communication technologies, it requires specialized knowledge to use [35]. Also, as the theory is employed and extended by researchers [4,16,27,36], its limitations are better understood, as well as new assumptions needed for applying it to the latest technologies [35]. For instance, researchers have extended Clark and Brennan's framework to include more affordances, such as tangibility [27], that did not exist when the theory of communication grounding was first established. In another work, Brennan et al. [5] introduced a data view mapping mechanism to reconcile two contradicting needs of collaborators in a multi-display system: the need to establish common ground among collaborators and the need to have private views to manipulate data.

Our work continues these efforts by the HCI and CSCW communities by extending Clark and Brennan's original framework to a more accessible visual tool that can be used by researchers and designers to understand and communicate how the design of specific user interface elements impacts grounding costs in different collaborative scenarios. We also leverage Clark's [9] concept of joint action ladders, which describe how communication between two people progresses. Joint action ladders capture momentary collective actions, in which a participant gets their partner to notice, perceive, and understand a verbal or non-verbal message, and finally consider responding. Each joint action ladder increments the group's common ground with new content. In this paper, we refer to 'joint action ladders' as 'joint actions' for simplicity. Clark [9] uses the concept to describe how face-to-face verbal communication takes place. Our visual framework adapts the definition of joint actions in the context of CSCW, and thus includes interactions with technology.

3 Joint Action Storyboards

We use the concept of joint actions to examine interactions with CSCW systems using a storyboard format, depicting an *instance of interest* as the unit of analysis. We define an instance of interest as a brief yet frequent interaction between collaborators and the CSCW technology. Instances may be identified by examining existing data, observed breakdowns in communication, or by creating hypothetical use case scenarios for the CSCW technology under study. Each frame in the storyboard represents a joint action and the corresponding grounding costs incurred. In a joint action, one executes a verbal or non-verbal action for a partner's attention (Figure 1). Meanwhile, the partner is noticing, perceiving, understanding the action, and at last considering a response.

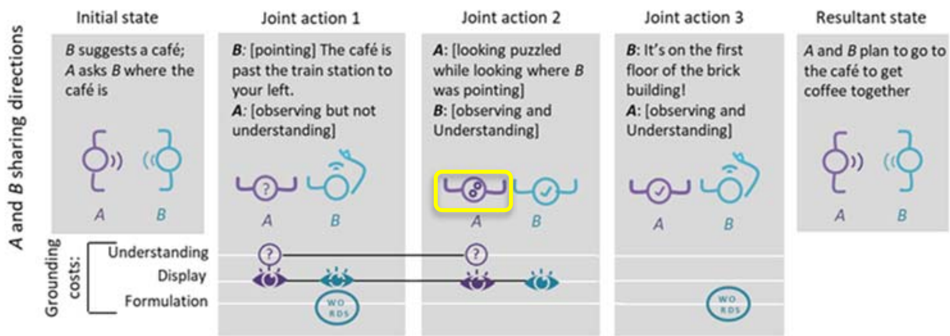


Figure 1. JASs depicting an instance of interest in which two people share directions to a cafe

This unit of analysis marks a fundamental difference between JASs and previous frameworks. Many existing frameworks, like the mechanics of collaboration [39] or Groupware Task Analysis [45] focus on high-level concepts and categorizations with respect to the entire collaborative activity. Our use of joint actions [9], however, enables fine-grained analysis of interaction with the system that may hinder or support the process of communication grounding.

Notably, by combining these different perspectives into a single framework, JASs consider multiple levels of analysis, e.g. individual and group levels [8,14,34]. At the group level, JASs depict verbal and physical interactions with teammates. At the individual level, the framework shows the collaborator's mental actions. Physical interactions with the CSCW system correspond to both individual and group levels.

In summary, the JASs framework is a qualitative analysis and communication tool for HCI researchers and technology designers that focuses on interaction minutiae. They can be used to analyze the impact of technology on communication grounding, and also provide a tool and vocabulary to articulate how subtle design decisions could impact communication grounding. When examining hypothetical cases, JASs also serve as a discount evaluation method to flag opportunities for revising the system's design..

4 How to Create JASs

JASs consist of two main parts that illustrate the collaborators' joint actions and the grounding costs incurred as they communicate (Figure 1). The top part of JASs depicts collaborators' initial states, followed by a sequence of joint actions, and finally the resultant state. The bottom part of JASs shows grounding costs incurred due to the specific user interface elements of the CSCW system. Each cost points to a potential opportunity for researchers to improve the user interface design to more effectively support communication grounding.

The development of JASs comprises four steps:

1. Choose an instance of interest
2. Break down the instance to joint actions
3. Create a storyboard of the joint actions
4. Mark grounding costs

In the following subsections, we will walk through the above steps by applying the framework to an introductory example of two people sharing directions to a café.

4.1 Choose an instance of interest

The first step in creating JASs is deciding which instance of interest is to be visualized. A researcher or technology designer needs to identify patterns of collaborative behaviour by either creating hypothetical scenarios or gathering and qualitatively analyzing data using existing techniques such as thematic analysis [3]. Gathering data can be done by running observations, field studies, or laboratory studies. Expected use cases of the system or uses of specific features can be used to generate hypothetical scenarios. An instance of interest shows a case where technology seems to support or get in the way of collaboration. In identifying such cases, it is useful to keep in mind the communication and coordination mechanisms [39], e.g. verbal/non-verbal messages or data transfer, and collaborative work styles [24,38,44] that are important in the progression of the collaborative task. Since a given instance is to be analyzed according to the exact series of actions by the participants, the instances should be short-lived, i.e. seconds long. Longer instances may not be representative of common behaviour in the CSCW system and may be broken down to shorter connected instances.

An instance of interest should represent an interaction that happens or is expected to happen often during collaboration, and thus, impacts grounding costs during the collaborative task. To identify such instances, we advise looking at prolonged interactions with the system (e.g., data sharing or manipulation during joint discussions), as well as frequent brief interactions (e.g., notification mechanisms):

1. *prolonged interactions* (e.g., where collaborators would be expected to be interacting with or discussing content for some minutes / hours) to ensure that access to and perception of the needed informational content, or even others' interactions with relevant task processes, does not incur significant grounding costs, and
2. *frequent brief interactions* (e.g., short-lived (milliseconds or seconds) group or system interactions) to ensure that the interface or interaction design aspects supporting those interactions do not introduce undue grounding costs that could accumulate over time and impact the broader flow of communication grounding process and collaboration in general.

Our instance of interest in the introductory example of two people sharing directions to a café is when one participant points to and describes the location of the coffee shop with respect to the surrounding buildings.

4.2 Break down the instance to joint actions

In this step, the instance of interest is broken down into the verbal and non-verbal exchanges between the participants and their (possibly concurrent) physical interactions with the environment (Figure 2). An example of a joint action is one telling their partner "The café is past the train station to your left." Meanwhile the partner is listening and understanding the message. We refer to the two people involved in the joint action as the initiator and the recipient. Importantly, in a given joint action during communication, a person may be the initiator or the recipient depending on who executes an action for the other's attention.

Physical non-verbal action

Verbal action

	Initial state	Joint actions 1	2	3	Resultant state
Scripts	*B* suggests a café; *A* asks *B* where the café is	***B***: [pointing] The café is past the train station to your left. ***A***: [observing but not understanding]	***A***: [looking puzzled while looking where *B* was pointing] ***B***: [observing and understanding]	***B***: It's on the first floor of the brick building! ***A***: [observing and understanding]	*A* and *B* plan to go to the café to get coffee together

Mental state

Figure 2. Joint actions of participants *A* and *B* while sharing directions.

Clark [9] describes mental actions of conversational participants during their communications. In any joint action, the initiator must get the recipient to notice, perceive, and understand the message, and finally consider executing a response [9]. At this point, the group's common ground is updated with new information. This new information includes knowledge of the communication that just took place. The recipient then assumes the role of the initiator and proceeds with a verbal or non-verbal response, which is yet another joint action and further increments common ground. If the recipient does not notice, perceive, or understand an action as expected by the initiator, the action is still joint but broken down. Including such failed joint actions in the analysis provides an opportunity to study where the user interface elements may hinder communication grounding.

In the above example, the recipient becomes the initiator and responds by, for example, looking puzzled while looking where their partner was pointing. As Clark [9] describes, the recipient of a joint action may only be in one of the following mental states with regards to the message. Note that the mental state of the recipient then shapes what happens in the next joint action:

1. Observing and understanding
2. Observing but not understanding
3. Not perceiving
4. Not noticing

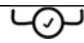
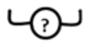
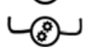
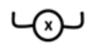
Figure 2 shows the joint actions for our example of two people, *A* and *B,* navigating a nearby café without a shared map. Note that there is a joint action each time one participant says or does something. The reaction of the recipient then follows as another joint action. It is important to include verbal, physical non-verbal, and mental actions of the participants (Figure 2) in the script as they are instrumental in creating the visualizations in the next step. The initial and resultant states are also included to provide context to the story.

Participants' representative labels (*A* and *B*) are highlighted in the scripts. The initiator's corresponding script appears first in the joint action. This ordering is a verification tool to make sure collaborators' responses to each other are captured in the sequence of joint actions [9]. Notably, the two conversational parties normally alternate the roles of initiator and recipient in subsequent joint actions. Even if *A* does not notice *B's* speech or their interaction with the environment, the subsequent joint action shows *A* is still doing whatever they were engaged in previously without any reaction. Meanwhile, *B* observes and understands that they were not noticed by *A*. In collaboration around technology, collaborators often transition between (sometimes brief) periods of loosely coupled and tightly coupled work. In this case, the participants may keep their roles as initiators and recipients in subsequent joint actions. Also, an interaction technique may require multiple steps to be performed immediately one after the other. . In such cases, collaborators keep their initiator and recipient roles in some subsequent joint actions until the recipient executes a response for the other.

4.3 Create a storyboard of the joint actions

We create storyboards in this step to visually capture the interactions between the participants and their environment, and the participants' mental states (Table 1) as those interactions take place. Depicting this contextually rich story of the joint actions generates empathy [32] and provides a tool to illustrate and understand how the participants adapt their collaborative processes to the specific user interface features of the CSCW system. Key entities that must be included are the participants and parts of environment they interact with.

Table 1. Mental states of participants. The view angle of the head icons may be changed as necessary.

Mental state	Icon	Mental state	Icon
Observing and understanding		Not perceiving	
Observing but not understanding	 	Not noticing	

The scripts generated in the previous step are noted above the images to provide a narration for the storyboards. Also, grey backgrounds group visualizations and scripts of individual joint actions. The head icons represent the two participants, *A* and *B*, and their corresponding mental states (Figure 1). Table 1 shows a sample of icons representing the mental action of the participants. Depending on the context under study, other icons may be used for a given mental action, for instance an icon depicting 'not seeing' or 'not feeling the haptic feedback' for the 'not perceiving' mental action. A concise label describing the instance of interest is displayed at the left side of the Initial state to make the illustration more self-descriptive. Participants' representative icons are distinguished by colour in the visualizations. The labels representing each participant, in our case *A* and *B*, are placed below the corresponding parties' images in addition to the colours to aid the readability and salience of the visualizations (Figure 1).

4.4 Mark grounding costs

In this step, grounding costs incurred at each joint action are marked below the corresponding participant's icon. We used the eleven communication grounding costs introduced by Clark and Brennan [10]. However, these costs were primarily based on using a medium to support conversation only, like a letter or telephone. Thus, we adapted some grounding costs' definitions, i.e. Production, Reception, Understanding, Delay, and Asynchrony (Table 2), for use in modern collaborative environments. For instance, collaborative software is the means of not only communication between team members, but also accessing material and functions that are essential to performing the task (e.g., creating a joint artefact). Grounding costs are incurred by the initiator, the recipient, or both parties (Table 2).

Table 2. Grounding costs and their definitions

Icon	Grounding Cost	Definition	Incurred by
	Formulation	Time and effort needed to formulate utterances	Initiator
	Reception	Costs of receiving a message, e.g. due to having to interact with the system	Recipient
	Production	Cost of producing a message, e.g. by typing or interacting with the system to retrieve data being referred to	Initiator
	Understanding	Cost of understanding a message, e.g. due to lack of contextual cues	Recipient
	Start-up	Costs of getting someone to notice a conversation has started	Both
	Display	Costs of presenting a gesture or an object, e.g. a piece of data to collaborators	Both
	Fault	Costs of making a mistake during communication	Both
	Repair	Costs of recovering from a mistake	Both
	Delay	Costs paid due to pauses during collaboration, e.g., loss of audio, or due to a need to pause one's ongoing work, e.g. to understand a screen update or to communicate with a partner about that	Initiator, Recipient, or Both
	Asynchrony	Incurred when a message is produced and received at different times, or in CSCW incurred when people transition between loosely coupled and tightly coupled work	Both
	Speaker-change	Costs needed to change the speaker to let the other party have a turn. These costs are generally low in synchronous CSCW	Both

Notably, these costs are not independent of each other and multiple costs may be incurred by a participant in a given joint action [10]. Lowering one cost may cause other costs to rise [4,10]. For instance, in our example, *B* could pay higher Formulation costs to lower Understanding costs for their partner *A*. It is important to mark all relevant grounding costs incurred due to the technology to provide a holistic view of the impact of user interface elements on communication grounding, and how the interface may be redesigned to counter or alleviate the costs. For a more elaborate list of costs' definitions, see the Appendix.

When identifying grounding costs, one needs to include costs that incur only due to the user interface elements of the system. This is to ensure a focused analysis of how the specific features of the CSCW environment facilitate or hinder communication grounding, and potentially identify opportunities to improve the user interface. For example, an initiator always needs to get the recipient's attention. If the two partners are co-located and can talk or if they are already engaged in tightly coupled work, then start-up costs are negligible and should not be included in the JASs. Also, in face-to-face communication, grounding costs for verbal dialogue exchange are minimal and are not considered in the analysis of the user interface features.

We recommend starting with the first joint action in the storyboard and scanning the grounding costs in Table 2 to see which are incurred by the collaborators. As demonstrated in Figure 1, draw a straight line below the visualizations for every identified cost and label the line accordingly. Then, recolour the corresponding grounding cost icon to match the image representing the participant who pays the cost and insert the icon on the line below the participant image. A grounding cost appears in subsequent joint actions until it is countered. *Sustained costs* persist across more than one join action, depicted by a back line connecting the respective cost icons. *Momentary costs*, however, appear in a single joint action at a time. Although joint actions in our storyboards do not necessarily last for the same amount of time, the distinction between sustained and momentary costs provides a notion of the amount of impact on grounding, depicted by the number of joint actions the costs span.

In our introductory example, sustained Understanding costs are incurred by *A* as they try to understand the directions with respect to the surrounding area (Figure 1). These costs persist until *B* clarifies the address they gave earlier. Participant *B* pays formulation costs when there is no map for reference as they need to describe the directions to the café. Additionally, sustained Display costs are noted for both participants as *B* tries to point to and show the location of the building to *A*. Those costs are countered when *B* specifies the characteristics of the target building.

5 Case Study: Analysis of Cross-Device Interface Designs

We were first motivated to explore visualizations of communication grounding when performing an analysis of Homaeian et al. [23]. Their system supported collaborative sensemaking between two people tasked with assessing the feasibility of shipping routes within a given geographic region using a multi-device environment comprised of personal tablet computers and a shared tabletop display. The tabletop showed a map of the region that was overlaid with data icons and displayed two bounding boxes around regions of interest (ROIs) that represented each of the two collaborators' tablets. By moving their ROI over a data icon, one would update their tablet's view to show only detailed data associated with that geographical region.

Homaeian et al. [23] studied how two interaction techniques, TOUCH and TILT, influenced collaboration in this setting. They reported that TOUCH and TILT impacted group work during a specific type of collaborative behaviour: showing a piece of data to a partner. When investigating that behaviour, we found that current methods of studying common ground, like conversational analysis and questionnaires, do not consider interaction with technology or the collaborators' mental actions. We wished to understand the grounding costs incurred by each technique to help us to understand the role that TOUCH and TILT played in facilitating collaboration. Since we were interested in the impact of technology on communication grounding, a process shaped by what goes on the participants' minds as well as their interactions with each other and their environment, it was crucial to include people's mental and physical steps in our analysis.

Although Homaeian et al. [23] reported that TOUCH and TILT supported joint and independent work periods differently, their analysis did not communicate how specific design choices led to that difference. Our visual framework enabled us to not only analyze the different grounding costs incurred by TOUCH and TILT, but also to *articulate* how interface design choices influenced them. JASs, by adapting the concept of joint actions [9] from the theory of communication grounding, allowed us to take into consideration the participants' mental actions and physical interactions with the environment as they used TOUCH and TILT to increment their common ground. We now show how JASs helped us conduct these analyses for the collaborative behaviour of showing data to a partner.

5.1 Showing data to a partner

We were initially motivated by difficulties in understanding how Homaeian et al.'s [23] participants shared data using the TOUCH and TILT interfaces. Notably, the tabletop display did not distinguish between users. Therefore, with TOUCH, participants often moved their partner's ROI to assist them in viewing data. With

TILT, however, this was not possible, and each participant always moved their own ROI. While subtle, these differences felt like they had a substantial impact on how the groups performed their tasks.

We found that existing methods like conversational analysis did not help in understanding these scenarios, since they do not consider people's interactions with the environment. Minutiae like collaborators' mental actions and positions around the shared tabletop, the orientation and location of data being discussed, and where individuals were looking while their partners interacted with the system had substantial impacts on communication grounding. Yet, existing analysis techniques tend to focus some of these factors and did not provide a satisfactory way of linking all of them.

We developed JASs for scenarios when one participant wishes to show a piece of data to their partner during tightly coupled work for each of the TOUCH and TILT interfaces (Figure **3** Top and Bottom, respectively). Our JASs show that the TILT interface required more collective effort, in the form of mental actions and interactions with the environment, than the TOUCH interface. In particular, two differences stand out based on our analysis. First, with the TILT interface, both participants incur sustained Display costs. Second, with TILT, Participant B first needs to retrieve and view the data on their personal tablet to facilitate an effective discussion, resulting in Reception costs that are not present when using the TOUCH interface (i.e., Figure **3** Top, Joint action #1).

Figure 3. Applying JASs reveals that the TOUCH interface (Top) facilitated more effective communication grounding compared to the TILT interface (Bottom).

5.2 Breakdowns in data sharing

While we found that the TOUCH interface afforded more effective grounding, we also found instances where it caused confusion when sharing information (Figure 4). Specifically, there were cases where one collaborator would move the others' ROI while they were focused on their personal tablet, surprising them. In these scenarios, there was no verbal communication, and so conversational analysis was not useful for understanding the impact on grounding. Even though conversation was absent, these groups were engaged in communication grounding, and we wanted to be able to articulate these costs and understand how they were linked to interface design choices. Note that in Joint actions #2 and #3 in Figure 4, the participants kept their roles as the initiator and recipient, unlike the pattern in most JASs storyboards, where the participants swap roles each frame of the JASs. This pattern is broken in this case because of the brief loosely coupled work period in which the collaborators were focused on their personal devices to understand a screen update, or lack thereof.

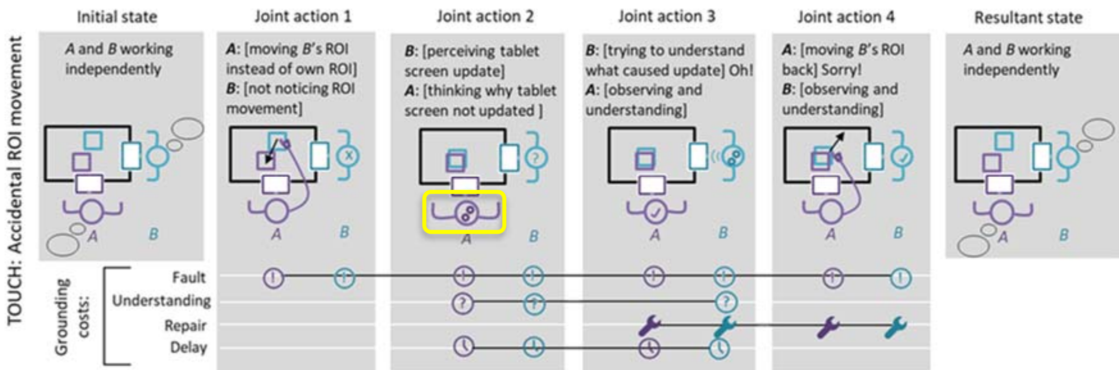


Figure 4. Using JASs to articulate breakdowns in grounding: even though TOUCH was more effective for grounding, groups paid several grounding costs when mistakes happened.

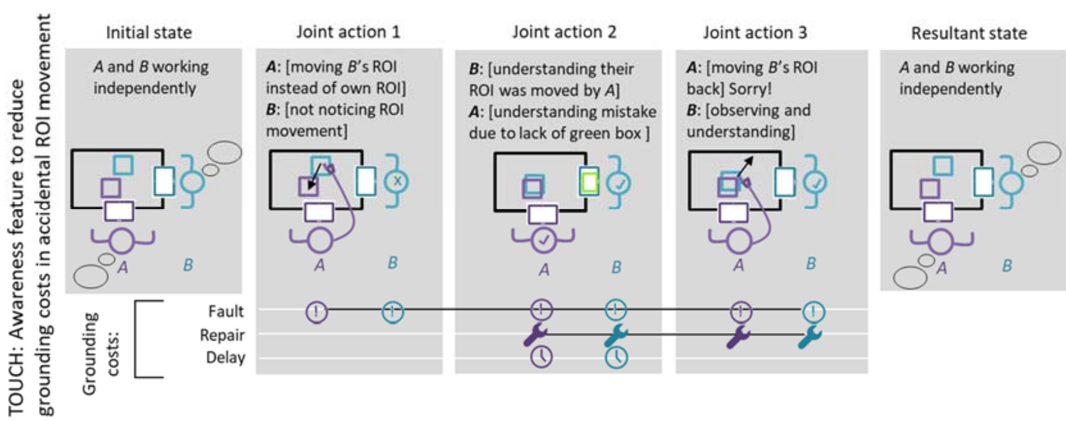


Figure 5. Using JASs to show a potential design alternative: highlighting one's tablet display to associate screen update with ROI movement could alleviate some grounding costs in case of accidental ROI movements.

In these cases, JASs serve as a useful tool to broadly think about the different grounding costs that may influence an instance, and their relationship to the user interface and physical workspace. They provide a systematic method of assessing grounding costs, accessible to novice and expert evaluators that can directly inform the design of complex, collaborative systems. For instance, in the scenario described in Figure 4, our analysis suggests that the user interface could be redesigned to inform the owner of a tablet once their screen is updated due to the movement of the associated ROI. As the Joint action #2 in Figure 5 depicts, the border of the tablet screen could flash to provide awareness of the ROI movement and the subsequent screen update. This design decision potentially alleviates sustained Understanding and Delay costs (Figure 4 Joint actions #2 and #3) and helps the group recover from the mistake faster (fewer joint actions in Figure 5).

6 Applying JASs to Other CSCW Systems

To demonstrate the key benefits of our framework, we next examine communication grounding in three case studies of techniques from the CSCW literature and a commercial tool. First, we show how our framework can be used as a discount evaluation method to analyze common ground and identify areas for improvement in the context of Marquardt et al.'s Tilt-to-Preview technique [31]. Second, we show how JASs can be used to depict physicality and the spatial environment for two digital tabletop sharing techniques. Finally, we show how our framework can be used to understand the impact of design choices on grounding during remote communication through an analysis of communication breakdown within a video conferencing system.

6.1 Discount evaluation tool for common ground

JASs are particularly useful as a discount evaluation method to understand the impact of specific user interface elements on grounding costs. To demonstrate this process, we performed an analysis of Marquardt et al.'s [31] GroupTogether, a system to explore cross-device interaction techniques for "micro-mobility". One technique in that system, called Tilt-to-Preview, supports tablet-to-tablet data sharing between two people standing beside each other. When a collaborator wishes to share an image, they tilt their tablet towards their partner's tablet, and touch the content they wish to transfer. A tinted edge then appears on their partner's tablet, together with a transient copy of the image that partially covers the receiver's screen. The receiver may touch the image to keep a permanent copy. In a preliminary user study, this cross-device transfer technique was found to be effortful by participants due to the added weight of sensors installed on the tablets [31]. However, our analysis also shows that its design requires sustained Production costs.

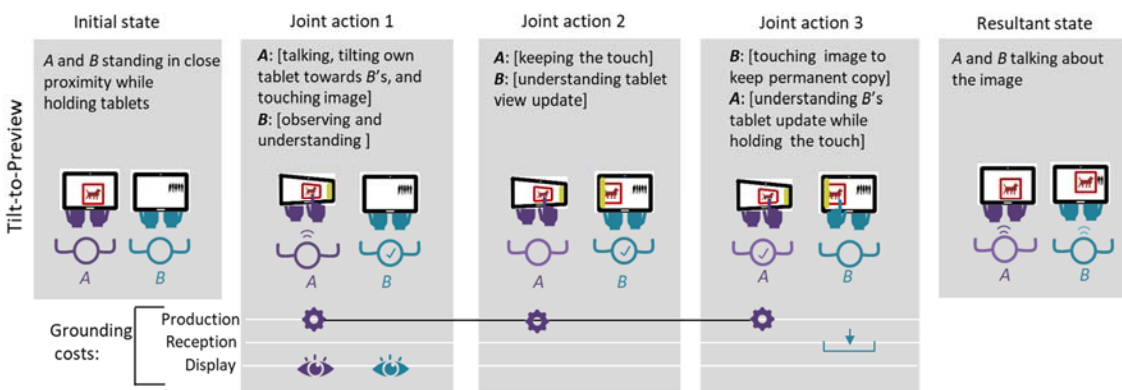


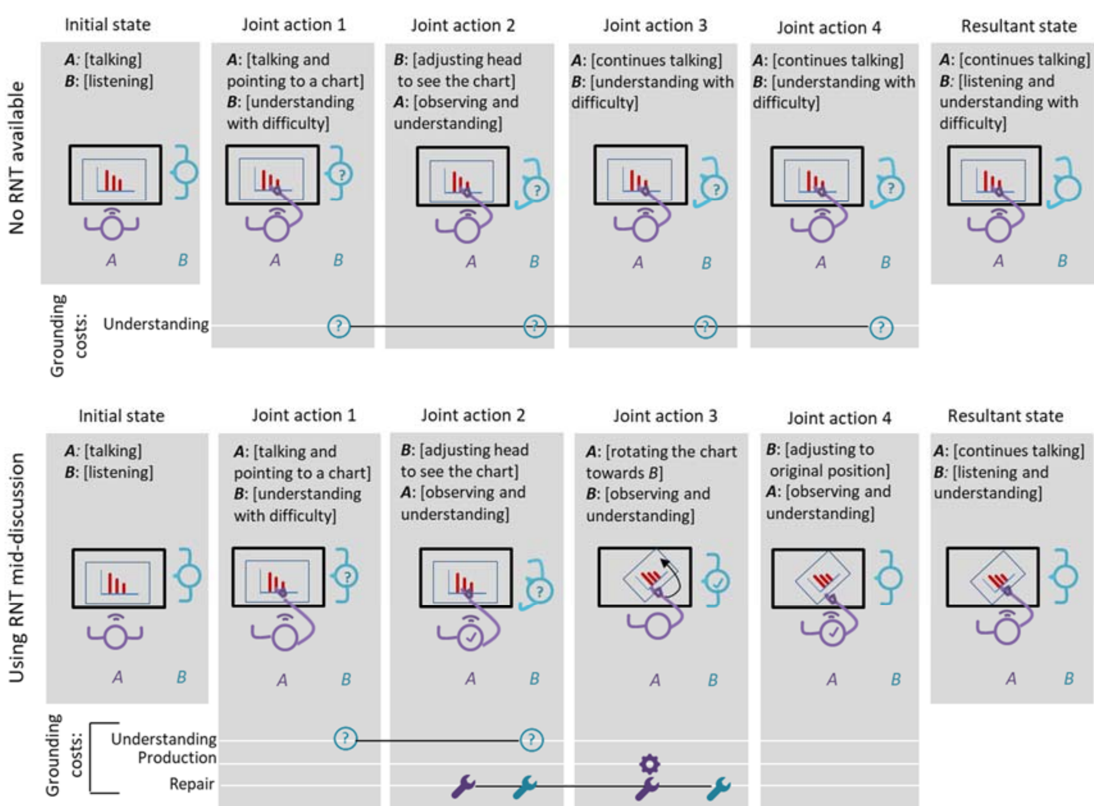
Figure 6. Applying JASs as a discount evaluation reveals sustained Production costs in data sharing with Tilt-to-Preview [31].

The JASs produced for this expected use case focus on the interaction minutiae of the required steps to complete the transfer of digital content from one device to another using the Tilt-to-Preview technique (Figure 6). Note that *A* keeps their role as the initiator in Joint actions #1 and #2 as the technique requires two steps to transfer the image: 1) tilting tablet and touching image (which automatically tints the edge of *A*'s tablet), and 2) sustaining tilt and touch actions (which automatically tints the edge to *B*'s tablet also). The JASs highlight the sustained Production costs that people incur every time they share content using this technique. This insight is articulated by depicting the joint actions of the collaborators while employing the Tilt-to-Preview technique and the solid black line pointing to sustained Production costs spanning the entire instance of interest (e.g., Production costs at the bottom of Joint actions #1 to #3 in Figure 6). These identified Production costs suggest an opportunity for design improvement.

While the JASs do not explicitly capture the magnitude of the grounding costs, they identify specific interaction and interaction sequences that could be considered for redesign. For instance, in this example, is it really necessary for the sender (*A* in Figure 6) to sustain both a device tilt AND a finger touch on the content item until the recipient (*B* in Figure 6) accepts the content? Would a less effortful design achieve the same goal and minimize Production costs? Alternatively, removing the requirement for the sender to keep touching the content would eliminate the more effortful of the two tilt / touch-while-holding-the-device-steady actions and reduce Production costs. Another interesting instance of interest not depicted in Figure 6, is when *B* is working independently and not anticipating a sudden change in their tablet view. A JASs analysis would show Understanding costs for *B* as they try to understand why unexpected content appeared on their screen (not currently included in Figure 6 because both parties are anticipating the transfer). In an alternative interface design, the tinted edge on the receiver's screen could appear shortly ahead of the transient copy of the shared item to help counter the Understanding costs.

6.2 Capturing physicality in interactions

JASs' visual format depicts physical interactions between collaborators and a CSCW system that are often difficult to understand and articulate using traditional analysis techniques. For example, rotation and translation (RNT) techniques for digital tabletops foster communication, coordination of actions, and comprehension of material during collaboration around tabletops [22,28,29,47]. But traditional analysis techniques often fail to articulate the connections between physical and mental actions and features of the system. In these cases, JASs provide a comprehensive method to understand and articulate how interface design can influence collaborative process.



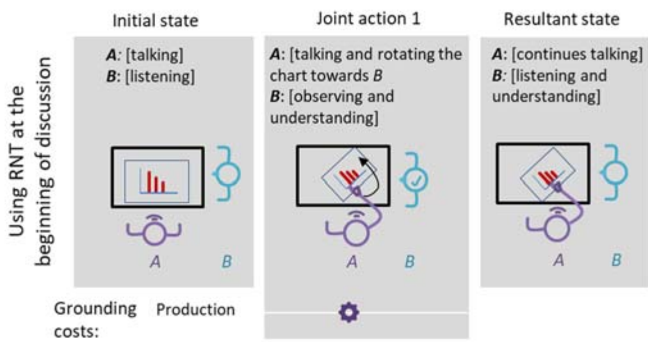


Figure 7. Illustrating physicality of CSCW technology and its importance during communication grounding. Top: Without RNT, *B* incurs sustained Understanding costs during the entire discussion. Middle: Using the RNT technique, *A* eliminates further understanding costs for *B*. Bottom: Using the RNT technique at the beginning of a discussion eliminates several grounding costs.

To better understand and articulate these interactions, we developed three JASs (Figure 7) that show different scenarios in which content is shared on a tabletop display: without RNT, when RNT is used to repair a breakdown in communication grounding, and when RNT is used to prevent breakdowns in communication grounding. In each scenario, the storyboard format of JASs captures both the physicality of sharing content around a tabletop display, and the associated mental actions and communication grounding costs.

When discussing materials displayed on the tabletop, without RNT, the recipient needs to either re-orient their head to the shared content, or do so mentally, to comprehend what is being shared (Figure 7 Top). Either of these activities would make the grounding process more effortful for the group. When the sender repairs a communication breakdown with RNT, they rotate the image to assist the recipient (Figure 7 Middle). Comparing the two JASs in Figure 7 Top and Middle reveals that while Participant *A* incurs Production and Repair costs, by doing so they eliminate sustained Understanding costs for their partner. This insight is depicted by the absence of grounding costs in Joint action #4 in Figure 7 Middle compared to Joint action #4 in Figure 7 Top. Finally, when RNT is available and the sender anticipates their collaborator's needs, only Production costs are incurred (Figure 7 Bottom).

6.3 Beyond collocated collaboration: Skype[1]

While we were motivated to develop JASs by our research on technologies that support synchronous, collocated collaboration, they are also applicable to a wide range of CSCW applications, including synchronous remote work. To demonstrate how JASs can be applied in these contexts, we analyzed a familiar use case where two people are conversing remotely over the commercial video conferencing tool Microsoft Skype[1] (Figure 8). Our analysis communicates the impact of the respective user interface elements on communication grounding and helps identify opportunities to improve the Skype interface to better support grounding.

We analyzed an instance of interest where the Wi-Fi signal suddenly becomes weak and audio and video jitter is introduced (Figure 8). The speaker may not notice the situation immediately and continue talking, leading to sustained Production costs (Joint actions #1-3, Figure 8 Top), as they continue talking until it becomes clear that there is a connection issue (Joint action #4, Figure 8 Top). Meanwhile, the listener pays Understanding costs as they try to see if the connection has been lost or not (Joint action #2, Figure 8 Top). The speaker will have to repeat the information they had conveyed during the connection failure once it is re-established. In the situation where the speaker is discussing content from a shared screen (e.g., sharing a PowerPoint presentation with their remote partner, as is common in virtual meetings), it may take some time

[1] https://www.skype.com/

for the speaker to notice the connection issues. Both the speaker and listener pay sustained Delay and Repair costs as their communication is interrupted (Delay: all Joint actions, Figure 8 Top; Repair: Joint actions #4-8, Figure 8 Top). Finally, if the listener decides to hang up and call back, the speaker incurs Understanding costs (Joint actions #4-5, Figure 8 Top), as they are left wondering what happened.

As much as possible, these tools should visualize any detected internet connectivity issues, and also status issues of connected parties to help minimize Understanding costs. For instance, in the commercial video conferencing tool Zoom[2], when a remote party's internet connection is poor, a red signal icon on their video / user box is displayed along with the text *"<user name>'s network bandwidth is low."*. Figure 8 Bottom shows the JASs for a hypothetical redesign of Skype that adds a similar connectivity awareness feature that makes the speaker aware of their partner's poor internet connection status (Joint action #1, Figure 8 Bottom). As the JASs shows, this allows the speaker to proactively initiate a repair, and eliminates Understanding costs for them. These process changes enabled by interfaces modifications that provide more awareness of the situation ultimately make the grounding process less effortful for the group.

[2] http://zoom.us

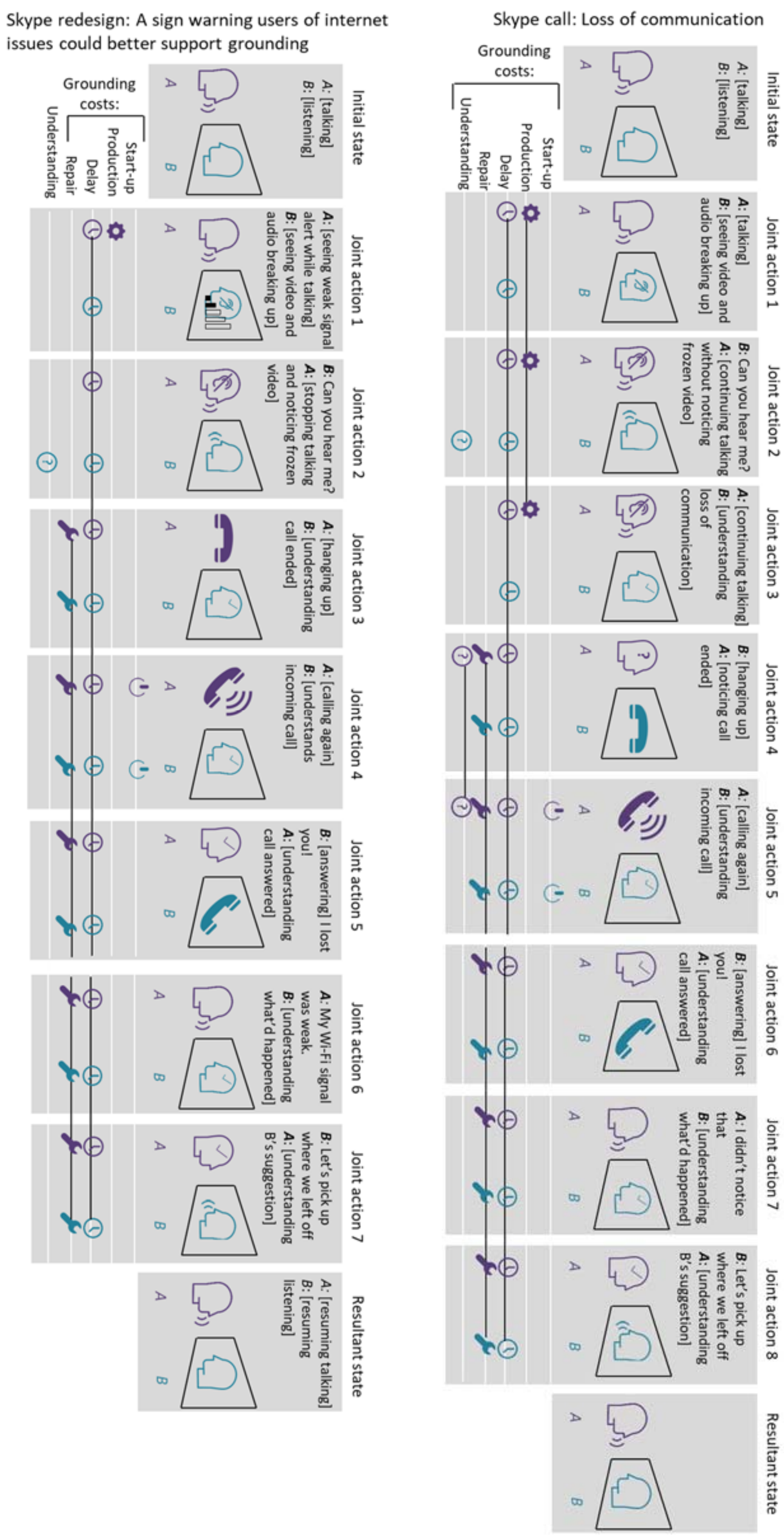


Figure 8. Top: JASs analysis of interruption during a remote synchronous collaborative scenario (Skype call). Bottom: Applying JASs to articulate a design alterative to alleviate some grounding costs.

7 Initial Experiences in Applying JASs in Practice

To better understand the level of expertise required to apply the JASs analysis technique to a CSCW use case, we ran a pilot study in the form a workshop with 4 HCI faculty and graduate students in our research institute. In a 90-minute session, we first gave a tutorial of how the JASs framework works by walking through the example of sharing directions to a café (Figure 1). Workshop participants were then asked to create JASs for the Tilt-to-Preview [31] use case described in Section 6.1.

We found that participants were able to identify some of the joint actions. However, there was some uncertainty around how to deal with overlap between costs. There was also some uncertainty about how many, or which, mental states could be represented for the recipient in a given joint action (i.e., only the following four: not noticing, not perceiving, not understanding, or observing and understanding). These informal experiences with the JASs tool showed that some level of expertise was needed to create JASs. Based on this finding, and comments from our workshop participants, we included a detailed description of the process of creating JASs in Sections 3 and 4 to help clarify the knowledge and the steps needed to create effective JASs representations.

8 Discussion

Our work was motivated by our own challenges in analyzing existing data, particularly in understanding how interfaces facilitate communication grounding, and finding that existing analysis techniques fell short. Clark and Brennan's [10] communication grounding framework provided a well-established foundation for HCI researchers and designers to draw on when examining communication effectiveness, yet it did not provide a sufficient means of linking communication effectiveness to CSCW tools. Like other CSCW researchers [5,12,20], we sought to leverage this strong foundation to help researchers and designers understand the impact of today's technologies on collaboration, and communication grounding specifically. By combining communication grounding with joint actions, JASs provide a novel method to link physical and mental interactions in CSCW environments, visualize and analyze interaction minutiae, and articulate design changes in complex and evolving CSCW settings.

JASs are a qualitative analysis tool. As with other qualitative analysis tools (e.g. heuristic evaluation [40], task analysis [40], and GOMS [13,40]), the goal of the analysis is not reaching consensus among researchers but to articulate design decisions and identify design opportunities. Agreement between individual researchers is not necessarily required, or even desirable, since "codes are the process not the product" [33:13]. Indeed, we believe the diversity and subjectivity of each researcher is valuable for capturing a broad range of perspectives and that a unique grounding cost or design opportunity flagged by one researcher does not discount aspects of the design identified by others, similar to unique usability issues identified by different heuristic evaluation inspectors. Similar to other analytic frameworks and tools, however, a skilled analyst will take some time to learn and apply the tool effectively. While we attempt to provide sufficient background on communication grounding and grounding costs as a stand-alone guide for creating JASs in this paper, as with any tool, the more practice, and contextual background the analyst has, the more insightful they will be at applying the tool.

As is true with other qualitative analytic tools, a researcher's existing knowledge about a CSCW system may influence what part of the data or system they pick for further examination using JASs. In this case, JASs allows for a fine-grained analysis of an instance of interest by linking physical and mental actions of the participants to the specific user interface elements. In fact, in our analysis of TOUCH and TILT [23], we were aware that the two techniques benefitted collaborative processes differently. What JASs provided was a tool and vocabulary to articulate how that difference was tied to the user interface elements, something that was not possible using the analysis methods employed in Homaeian et al.'s original work [23].

Our case studies demonstrate how JASs may be applied outside of traditional collocated, synchronous collaboration settings. While, space constraints do not allow an exhaustive list, we envision that they are applicable in a range of contexts, including augmented- or mixed-reality technologies.

8.1 Linking physical and mental interactions and grounding costs

JASs provide an analysis structure, vocabulary, and visual framework to help CSCW researchers and designers understand the impact of the system design on communication grounding. While researchers highly familiar with the grounding process and grounding costs may not need these tools, as researchers who have both taught and used these concepts in existing analysis approaches (e.g. conversational analysis, qualitative video coding), we recognize how complex and context-dependent the grounding process and costs are. We have found the use of specific instances of interest and the visual framework of the storyboard structure to be highly useful to help focus analytic efforts on key aspects of group interactions with CSCW technology.

A key advantage of JASs is that they convey an integrated view of the physical and mental interactions involved in the instance of interest, together with the relevant grounding costs, to provide an overall depiction of the situation that helps analysts understand and communicate the strengths and weaknesses of the studied interface or interaction design features. Indeed, through developing JASs for the presented case studies, we have revealed design opportunities to help minimize grounding costs. For instance, the JASs analysis of Tilt-to-Preview [31] revealed an opportunity to reduce sustained Production costs incurred with the current system design (Figure 6) and JASs analysis of Homaeian et al.'s [23] TOUCH interface revealed an opportunity to reduce Understanding costs that can arise in some situations with the current design (Figure 3 and Figure 4).

8.2 Focused analysis of interaction minutiae

CSCW research has shown that small design changes in a user interface design can impact collaboration [21]. Yet, traditional collaborative analysis methods like conversational analysis (i.e. word counting or word coding) or video analysis are insufficient for understanding the impact of a tool on collaboration, or for understanding what specific aspects of the technology should be improved. Analysis frameworks like JASs can provide the necessary structure, vocabulary, and concepts to help researchers focus on key aspects of a group's interaction directly or indirectly involving the technology, at an appropriate level of detail, to help uncover challenges that may be overlooked by all but the keenest analyst using other tools. For instance, during discussions around a piece of data using the TOUCH and TILT techniques [23], often people didn't verbally instruct a partner to move their ROI to the respective location on the map. This subtle behaviour would render current communication grounding techniques, such as conversational analysis, less effective. However, JASs captures how people adapted their grounding process to the unique affordances of TOUCH and TILT by looking at the details of people's interaction with the CSCW.

The use of "instances of interest" in our JASs approach was intentional. Clark and Brennan [10] and others [21] have shown how the minutiae of communication and collaboration processes around CSCW can impact the entire experience over time. For example, Gutwin and Greenberg [21] discuss that even a subtle change such as highlighting a button when pressed has an important role in providing awareness to the team about an action that a collaborator executed. Instances of interest focus on interaction minutiae; they enable a focused analysis of how specific user interface design choices might impact people's experiences around CSCW technology.

8.3 Articulating design changes in CSCW environments

JASs' storyboard format enables the creation of simple diagrams that depict the spatial layout and body orientation of people relative to each other and relative to the technology they are using, or even the environment in which they are working. Indeed, effective grounding involving CSCW systems depends heavily on group members' ability to notice and perceive both verbal and nonverbal communication from collaborators, as well as any displayed information (textual, visual, graphical, auditory, etc.) [10]. Deictic referencing (pointing, gesturing, etc.) is important for communication and grounding [2,10]. The multitude of form factors, devices, number of devices, and various interface and interaction design options available in today's technologies can introduce many challenges for effective non-verbal communication that can impact both the informational and integrational (i.e., procedural) aspects of communication [1]

Thus, for CSCW technologies that involve mobility or physical interaction or collocated collaboration, the spatiality of people and any technology they are using may be extremely important to examine when assessing how well a tool supports grounding. This is perhaps obvious for the use cases examined in our case studies that involved collocated CSCW technologies, like cross-device interactions or discussions of content on a digital tabletop. For instance, in the RNT Case Study (Figure 7), the orientation of collaborators to each other, and relative to the tabletop and its displayed content was essential for understanding the potential barriers to grounding that groups can face in that technology setting.

However, this may also be relevant to capture when examining collaboration involving augmented- or mixed-reality tools, where one or more collaborator is interacting with one's physical surroundings as part of the digital interactions. Moreover, even a remote video call may involve physicality in which spatial interactions are relevant. Many of us have experienced video calls where one person is walking around a space with a smartphone, tablet, or laptop, trying to show their remote partner(s) something in the physical space. This can often be very frustrating and disorienting for all parties. By capturing the relative spatial relationships of all parties and devices and relevant interfaces involved, JASs could help identify features of the available video chat tools that either hinder or support communication grounding.

9 Limitations

By design, JASs focus on the minutiae of collaborative interactions, and the impact that CSCW technology interfaces have on those interactions. Similarly, JASs are intended to specifically assess the impact of a technology on a group's communication processes. Due to this focus, we envision JASs being used in conjunction with other assessment tools and measures commonly used in CSCW research and practice (e.g., conversational analysis, qualitative video analysis, interaction log analysis), that focus on other aspects of the communication or collaboration

We designed the JASs visualizations to accommodate a wide range of technologies, and to be adaptable to new types of devices as they emerge. However, when using JASs as a "discount usability" technique for technology design, observations of group interaction with the system are unavailable to identify suitable instances of interest. In this case, expected use cases involving collaborative interactions with the technology will need to be generated first, and then specific instances of interest can be identified from those use cases. We believe that the exercise of generating these instances, and envisioning interaction with a new device or system are also beneficial activities in themselves.

Clark and Brennan, as well as others, have noted that their communication grounding framework was not exhaustive [10,16]. Indeed, additional grounding costs and communication medium (or technology) affordances have been proposed by Brennan and others in subsequent research. For example, Brennan et al. [4,6] defined additional grounding costs such as monitoring and face management, that refer to the cost of monitoring a partner's focus of attention or the objects they are manipulating within the environment, and the cost of maintaining politeness in a conversation, respectively. These costs are paid by either party. We did not include these costs in our JASs analyses, as they were generally consistent in the instances of interest under consideration. However, these or other additional relevant grounding costs could easily be considered in the analysis by adding extra cost lines at the bottom of each joint action frame in the storyboard.

Finally, JASs were designed to assess grounding costs for small-groups. Visualizing the mental and physical interactions for a large number of people may not be feasible using current representations. However, we could imagine scenarios where a large number of “audience” members, using the same or similar system designs (software and hardware) might be represented collectively as a single “audience”, “student”, or other . role in the JASs diagrams. We have even conducted thought experiments of applying this tool to help assess the many available video conferencing and virtual classroom tools being used for remote course delivery during the current COVID-19 pandemic to understand the grounding costs involved in these systems. Further work is needed to understand the feasibility of this approach.

10 Conclusion

We presented the development of JASs, a visual framework for analyzing and articulating the impact of specific user interface designs on communication grounding. We demonstrated JASs’ benefits over existing methods for HCI and CSCW research by applying the framework to a number of case studies of interaction techniques from the literature. We also discussed that by capturing collaborators’ mental and physical interactions with each other and the CSCW environment in in integrated view, JASs enable a focused analysis of interaction minutiae and provide a vocabulary and tool for researchers and technology designers to communicate the impact of user interface designs on communication grounding. Finally, we discussed the limitations of our method.

We believe that JASs can be useful in understanding interaction with novel devices and in novel settings. For instance, we are increasingly taking advantage of home assistants like Amazon Alexa and Google Home, in which communication grounding occurs between a human and an artificial intelligence (AI) collaborator. Similarly, collaborative augmented- and virtual-reality applications are an active area of HCI research, and undoubtedly need to support communication grounding in virtual, rather than physical spaces. The JASs visual framework is both powerful and flexible enough to articulate communication grounding in these use cases and is a promising avenue for future work.

11 Acknowledgments

This work is supported by the Natural Sciences and Engineering Research Council of Canada. We would like to thank our colleagues at the University of Waterloo and the University of Guelph and anonymous reviewers for providing invaluable feedback and comments on our work. Head icons in Figure 8 and Production, Delay, and Repair cost icons in Table 2 have been adapted using resources from Flaticon.com.

REFERENCES

[1] Michael Argyle. 1969. *Social Interaction*. Methuen, London.

[2] Mathilde M. Bekker, Judith S. Olson, and Gary M. Olson. 1995. Analysis of gestures in face-To-face design teams provides guidance for how to use groupware in design. In *Proceedings of the 1st Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques (DIS ’95)*, ACM, New York, NY, USA, 157–166. DOI:https://doi.org/10.1145/225434.225452

[3] Richard E. Boyatzis. 1998. *Transforming qualitative information : thematic analysis and code development*. Sage Publications. Retrieved July 3, 2018 from http://psycnet.apa.org/record/1998-08155-000

[4] Susan E. Brennan and Calion B. Lockridge. 2006. Computer-Mediated Communication: Cognitive Science Approach. In *Encyclopedia of Language & Linguistics*. Elsevier, 775–780. DOI:https://doi.org/10.1016/b0-08-044854-2/00861-0

[5] Susan E. Brennan, Klaus Mueller, Greg Zelinsky, I. V. Ramakrishnan, David S. Warren, and Arie Kaufman. 2006. Toward a Multi-Analyst, Collaborative Framework for Visual Analytics. In *Proceedings of the IEEE Symposium On Visual Analytics Science And Technology (VAST ’06)*, IEEE, 129–136. DOI:https://doi.org/10.1109/VAST.2006.261439

[6] Susan E. Brennan and Justina O. Ohaeri. 1999. Why do electronic conversations seem less polite? the costs and benefits of hedging. In *Proceedings of the International Joint Conference on Work Activities Coordination and Collaboration (WACC ’99)*, ACM, New York, NY, USA, 227–235. DOI:https://doi.org/10.1145/295665.295942

[7] Stuart K. Card, Thomas P. Moran, and Allen Newell. 1980. The keystroke-level model for user

performance time with interactive systems. *Commun. ACM* 23, 7 (1980), 396–410. DOI:https://doi.org/10.1145/358886.358895

[8] John M. Carroll, Gregorio Convertino, Mary Beth Rosson, and Craig H. Ganoe. 2008. Toward a conceptual model of common ground in teamwork. In *Macrocognition in Teams: Theories and Methodologies*. Ashgate Publishing Ltd, 87–105. DOI:https://doi.org/10.1201/9781315593166-6

[9] Herbert H Clark. 1996. *Using Language*. Cambridge University Press.

[10] Herbert H Clark and Susan E Brennan. 1991. Grounding in communication. In *Perspectives on socially shared cognition*, L B Resnick, J M Levine and S D Teasley (eds.). American Psychological Association, Washington, DC, US, 127–149. DOI:https://doi.org/10.1037/10096-006

[11] Gregorio Convertino, Helena M. Mentis, Mary Beth Rosson, John M. Carroll, Aleksandra Slavkovic, and Craig H. Ganoe. 2008. Articulating common ground in cooperative work: Content and process. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*, ACM, New York, NY, USA, 1637–1646. DOI:https://doi.org/10.1145/1357054.1357310

[12] Gregorio Convertino, Helena M. Mentis, Mary Beth Rosson, Aleksandra Slavkovic, and John M. Carroll. 2009. Supporting content and process common ground in computer-supported teamwork. In *Proceedings of the SIGCHI Conference on Human factors in Computing Systems (CHI '09*, ACM, New York, NY, USA, 2339–2348. DOI:https://doi.org/10.1145/1518701.1519059

[13] Alan Dix, Janet Finlay, Gregory Abowd, and Russell Beale. 2004. *Human-computer interaction*. Pearson Education UK.

[14] Yrjö Engeström. 1990. *Learning, working and imagining : twelve studies in activity theory*. Orienta-Konsultit Oy.

[15] Jean E. Fox Tree and Nathaniel B. Clark. 2013. Communicative Effectiveness of Written Versus Spoken Feedback. *Discourse Process.* 50, 5 (2013), 339–359. DOI:https://doi.org/10.1080/0163853X.2013.797241

[16] Jean E. Fox Tree, Sarah A. Mayer, and Teresa E. Betts. 2011. Grounding in Instant Messaging. *J. Educ. Comput. Res.* 45, 4 (2011), 455–475. DOI:https://doi.org/10.2190/EC.45.4.e

[17] Darren Gergle. 2017. Discourse Processing in Technology-Mediated Environments. In *The Routledge Handbook of Discourse Processes*. Routledge, 191–221.

[18] Darren Gergle, Robert E. Kraut, and Susan R. Fussell. 2004. Language efficiency and visual technology: Minimizing collaborative effort with visual information. *J. Lang. Soc. Psychol.* 23, 4 (2004), 491–517. DOI:https://doi.org/10.1177/0261927X04269589

[19] Darren Gergle, Robert E. Kraut, and Susan R. Fussell. 2004. Action as language in a shared visual space. In *Proceedings of the ACM conference on Computer supported cooperative work (CSCW '04)*, ACM, New York, NY, USA, 487–496. DOI:https://doi.org/10.1145/1031607.1031687

[20] Darren Gergle, Robert Kraut, and Susan Fussell. 2013. Using visual information for grounding and awareness in collaborative tasks. *Human–Computer Interact.* 28, 1 (2013), 1–39. DOI:https://doi.org/10.1080/07370024.2012.678246

[21] Carl Gutwin and Saul Greenberg. 1998. Design for individuals, design for groups: tradeoffs between power and workspace awareness. In *Proceedings of the 1998 ACM conference on Computer supported cooperative work (CSCW '98)*, ACM, Seattle, WA, USA, 207–216. DOI:https://doi.org/10.1145/289444.289495

[22] Mark S. Hancock, Frédéric D. Vernier, Daniel Wigdor, Sheelagh Carpendale, and Chia Shen. 2006. Rotation and translation mechanisms for tabletop interaction. In *Proceedings of the First IEEE International Workshop on Horizontal Interactive Human-Computer Systems (TABLETOP '06)*, 79–86. DOI:https://doi.org/10.1109/TABLETOP.2006.26

[23] Leila Homaeian, Nippun Goyal, James R. Wallace, and Stacey D. Scott. 2018. Group vs Individual: Impact of TOUCH and TILT Cross-Device Interactions on Mixed-Focus Collaboration. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*, ACM, New York, NY, USA, 73. DOI:https://doi.org/10.1145/3173574.3173647

[24] Petra Isenberg, Danyel Fisher, Sharoda A Paul, Meredith Ringel Morris, Kori Inkpen, and Mary Czerwinski. 2012. Co-Located Collaborative Visual Analytics around a Tabletop Display. *IEEE Trans. Vis. Comput. Graph.* 18, 5 (2012), 689–702. DOI:https://doi.org/10.1109/tvcg.2011.287

[25] David Kirk, Tom Rodden, and Danaë Stanton Fraser. 2007. Turn it this way: Grounding collaborative action with remote gestures. In *Proceedings of the SIGCHI Conference on Human*

Factors in Computing Systems (CHI '07), ACM, New York, NY, USA, 1039–1048. DOI:https://doi.org/10.1145/1240624.1240782

[26] Theodora Koulouri, Stanislao Lauria, and Robert D. Macredie. 2017. The influence of visual feedback and gender dynamics on performance, perception and communication strategies in CSCW. *Int. J. Hum. Comput. Stud.* 97, (2017), 162–181. DOI:https://doi.org/10.1016/j.ijhcs.2016.09.003

[27] Robert E. Kraut, Susan R. Fussell, Susan E. Brennan, and Jane Siegel. 2002. Understanding effects of proximity on collaboration: Implications for technologies to support remote collaborative work. In *Distributed work*. MIT Press, 137–162.

[28] Russell Kruger, Sheelagh Carpendale, Stacey D. Scott, and Saul Greenberg. 2004. Roles of orientation in tabletop collaboration: Comprehension, coordination and communication. *Comput. Support. Coop. Work* 13, 5–6 (2004), 501–537. DOI:https://doi.org/10.1007/s10606-004-5062-8

[29] Russell Kruger, Sheelagh Carpendale, Stacey D. Scott, and Anthony Tang. 2005. Fluid integration of rotation and translation. In *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '05)*, ACM, New York, NY, USA, 601. DOI:https://doi.org/10.1145/1054972.1055055

[30] John D. Lee, Alex. Kirlik, and Marvin J. Dainoff. 2013. *The oxford handbook of cognitive engineering*. Oxford University Press.

[31] Nicolai Marquardt, Ken Hinckley, and Saul Greenberg. 2012. Cross-device interaction via micro-mobility and F-formations. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology (UIST '12)*, 13–22. DOI:https://doi.org/10.1145/2380116.2380121

[32] Bella. Martin and Bruce M. Hanington. 2012. *Universal methods of design : 100 ways to research complex problems, develop innovative ideas, and design effective solutions*. Rockport Publishers.

[33] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and Inter-rater Reliability in Qualitative Research. *Proc. ACM Human-Computer Interact.* 3, CSCW (November 2019), 1–23. DOI:https://doi.org/10.1145/3359174

[34] Joseph Edward McGrath. 1984. *Groups: Interaction and performance*. Prentice-Hall, Inc., Eaglewood Cliffs, N.J.

[35] Andrew Monk. 2003. Common Ground in Electronically Mediated Communication: Clark's Theory of Language Use. In *HCI Models, Theories, and Frameworks: Toward a Multidisciplinary Science*. Elsevier Inc., 265–289. DOI:https://doi.org/10.1016/B978-155860808-5/50010-1

[36] Andrew Monk and Leon Watts. 2000. Peripheral participation in video-mediated communication. *Int. J. Hum. Comput. Stud.* 52, 5 (2000), 933–958. DOI:https://doi.org/10.1006/ijhc.1999.0359

[37] C.E. Nehme, Stacey D. Scott, M.L. Cummings, and C.Y. Furusho. 2006. Generating Requirements for Futuristic Heterogeneous Unmanned Systems. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 235–239.

[38] Thomas Neumayr, Hans Christian Jetter, Mirjam Augstein, Judith Friedl, and Thomas Luger. 2018. Domino: A descriptive framework for hybrid collaboration and coupling styles in partially distributed teams. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW (2018). DOI:https://doi.org/10.1145/3274397

[39] David Pinelle, Carl Gutwin, and Saul Greenberg. 2003. Task analysis for groupware usability evaluation: Modeling shared-workspace tasks with the mechanics of collaboration. *ACM Trans. Comput. Interact.* 10, 4 (2003), 281–311. DOI:https://doi.org/10.1145/966930.966932

[40] Jenny Preece, Yvonne Rogers, and Helen Sharp. 2007. *Interaction design : beyond human-computer interaction* (2nd ed.). John Wiley.

[41] Andrew D. Rice and Jonathan W. Lartigue. 2014. Touch-level model (TLM): evolving KLM-GOMS for touchscreen and mobile devices. In *Proceedings of the ACM Southeast regional conference (ACM SE '14)*, ACM, New York, NY, USA, 1–6. DOI:https://doi.org/10.1145/2638404.2638532

[42] Stacey D. Scott, T C Nicholas Graham, James R. Wallace, Mark Hancock, and Miguel Nacenta. 2015. "Local Remote" Collaboration: Applying Remote Group AwarenessTechniques to Co-Located Settings. In *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing (CSCW '15 companion)*, ACM, New York, NY, USA, 319–324. DOI:https://doi.org/10.1145/2685553.2685564

[43] Da Silva, Stacey D. Scott, and M.L. Cummings. 2007. *Design Methodology for Unmanned Aerial Vehicle (UAV) Team Coordination*.

[44] Anthony Tang, Melanie Tory, Barry Po, Petra Neumann, and Sheelagh Carpendale. 2006. Collaborative coupling over tabletop displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*, ACM, 1181–1190. DOI:https://doi.org/10.1145/1124772.1124950

[45] Gerrit C. Van Der Veer, Bert F. Lenting, and Bas A.J. Bergevoet. 1996. GTA: Groupware task analysis - Modeling complexity. *Acta Psychol. (Amst).* 91, 3 SPEC. ISS. (April 1996), 297–322. DOI:https://doi.org/10.1016/0001-6918(95)00065-8

[46] James R Wallace, Saba Oji, and Craig Anslow. 2017. Technologies, Methods, and Values: Changes in Empirical Research at CSCW 1990-2015. *Proc. ACM Hum.-Comput. Interact* 1, CSCW (November 2017), 1–18. DOI:https://doi.org/10.1145/3134741

[47] Daniel Wigdor, Chia Shen, Clifton Forlines, and Ravin Balakrishnan. 2006. Effects of display position and control space orientation on user preference and performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*, 309–318. DOI:https://doi.org/10.1145/1124772.1124819

APPENDIX

Detailed definitions of grounding costs [10] and, in case of Production, Reception, Understanding, Delay, and Asynchrony costs, how they have been adapted for CSCW environments:

-
- **Formulation costs** refer to the time and effort involved in composing a message. The cost depends on the complexity of the message, whether it includes familiar or unfamiliar content, and if the content needs to be flawless. Formulation costs are lower when the group shares a visual space [18,19].
- **Reception costs** refer to the effort that it takes to receive a message. They may include interacting with the system to view data a partner is referring to. If one has to wait a long time for the message to be produced or if the user interface requires significant interaction effort for the message to be accessed, the addressee pays a high cost.
- **Production costs** refer to the cost associated with interacting with technology to produce a message, for e.g., by typing. In CSCW, production costs may include interacting with the system to retrieve or share data that collaborators are discussing.
- **Understanding costs** refer to the effort involved in interpreting the received message, e.g. due to lack of contextual cues or complexity. If awareness features are not sufficient or up-to-date in the system, collaborators also pay understanding costs.
- **Startup costs** are the cost of getting a collaborator to notice that a partner has said something or is initiating communication. In co-located CSCW, startup costs are generally low as people can use verbal communication.
- **Display costs** are associated with using non-verbal communication, for e.g., nodding or pointing, or showing an object (e.g. a piece of data) to a collaborator.
- **Fault costs** are incurred when a partner makes a mistake during collaboration, for e.g., they send the wrong data item.
- **Repair costs** are the cost of recovering from a mistake, for e.g., understanding the situation and sending the right data.
- **Delay costs** are associated with pauses that occur during collaboration. In CSCW, partners may need to stop their ongoing work to, for e.g., make sense of an unanticipated screen update. In this case only the affected party incurs Delay costs. If they need to interrupt a partner to talk about the update, then both parties pay the cost.
- **Asynchrony costs** are paid when collaborators transition between loosely coupled work to tightly coupled work periods as they need to shift their focus to the new work style and a potentially different context.
- **Speaker-change costs** refer to the costs needed to change from one speaker to another. These costs are generally low in synchronous CSCW (focus of JASs), especially with audibility, as there are sufficient cues for turn taking.