

Making Time: Pausing to Coordinate Video Instructions and Practical Tasks

Sylvaine Tuncer 

KTH Royal Institute of Technology, Sweden

Oskar Lindwall

University of Gothenburg, Sweden

Barry Brown

Stockholm University, Sweden

Using video recordings as data to study how dyads follow instructional videos to achieve practical tasks, this article focuses on how participants coordinate the temporality of the video with that of their task by pausing the video. We examine three types of pausing, each displaying participants' online understanding of the instructions and different articulations between demonstrations and practical task: pausing to raise a correspondence problem, to keep up with the video, and to turn to action. From this exemplar case, we discuss how ordinary people experience and make time with interactive media.

Keywords: instructional videos, coordinating multiple temporalities, following instructions, making time, conversation-analytic study

INTRODUCTION

Video tutorials provide step-by-step instructions on ordinary tasks, such as using chopsticks, fixing a broken dishwasher or applying make-up. They have become a major pedagogical medium in the past decade, nowadays readily available online by the hundreds of millions. As such, they are also naturally available for research, in particular for media studies and educational research, and they have been studied with both quantitative, coding methods (Kruse and Veblen 2012), and qualitative, critical discourse analytical methods (Bhatia 2018; Dynel 2014; Frobenius 2011, 2013; García-Rapp 2017). Existing studies examine the content of the media, the economics of video production, how videos are constructed, and how user-created

Direct all correspondence to Sylvaine Tuncer, Department of Robotics, Perception and Learning, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden; e-mail: sylvaine@kth.se.

Symbolic Interaction, (2020), p. n/a, ISSN: 0195-6086 print/1533-8665 online.

© 2020 Society for the Study of Symbolic Interaction (SSSI)

DOI: 10.1002/SYMB.516

content has reconfigured the media landscape (Strangelove 2010). While it is possible to watch video tutorials just as entertainment, like cooking shows on broadcast TV, the original goal of instructional videos is to provide the know-how and enable people to do, fix, or learn something, and yet the above research leaves unexamined how instructional videos are actually used.

In this article, we aim to address this gap with an analysis of participants using instructional videos to achieve practical tasks. We recruited ten pairs of participants and gave them a specific task of which they had little or no previous experience, such as replacing a bicycle chain, picking a lock with paper clips, or using a specific make-up technique. We then asked them to find and use videos online to help carry out the task. By video recording them while they were following the instructional videos they had chosen, we documented how they achieved (or failed to achieve) the instructed task. Thus, with this case study, we present one of the few detailed analyses of how online videos are watched and used for practical purposes.

Our focus is on how, with this particular instructional medium, participants need to coordinate the different *temporalities* of what is demonstrated in the video and their attempts to follow the demonstrations (typically because the former progresses faster than the latter). Participants do this mainly by pausing the video, adjusting the delivery of the instructions to the progression of their task. Thus, our enquiry into instructional video, and more broadly interactive media use, led us to focus on how participants coordinate different temporalities, the methodical ways in which they pause before or while they attempt to reproduce the demonstrations. When participants pause an instructional video they display through their interactions an orientation to the relative progression of the task in the video and of their task in the physical world, as mutually constitutive. We focus on three types of pausing: pausing to raise a correspondence problem, pausing to keep up, and pausing to turn to action. In the empirical, analytic section, we illustrate and describe each type of pausing using the video data to give examples of how these pauses are achieved. In the light of these findings, we argue that the practices associated with these different types of pausing exhibit how instructions are analyzed by participants, something which involves their ongoing segmentation into separable parts, and the continuous establishment of a correspondence between the instruction world and the physical world. Lastly, we discuss how practices developed in the use of interactive technologies are critical to users and ordinary people's experience and management of time, and we explain how such activities can be seen as a form of collective action expanding beyond the local situation.

This case documents how, in our digitally mediated lives, we must continually weave together onscreen instructions with practical tasks, creating a coherent and seamless course of action out of the two. Watching and doing are inseparable: they support each other, they do not compete with one another. This study thus teases out the complex relationships between what is done with a device *and* around it.

Following Instructions

Instructions can take various forms, from text and images to assemble a piece of furniture (Garfinkel 2002) to graphical, interactive maps to find one's way (Brown and Laurier 2005; Laurier, Brown, and McGregor 2015). Instructions provide a frame of reference for (a course of) actions, but they are decontextualized renderings, by essence indexical, and necessarily incomplete. Whatever their form, it takes work to transform instructions into concrete actions. They are two-dimensional, whereas tools, materials, and so on, exist in a three-dimensional space, and each step of the task is likely to transform material configurations. This article builds on ethnomethodology's long tradition of studies on the lived work of following instructions, which investigate how members transform decontextualized descriptions into actions in the here-and-now (Garfinkel 2002; Livingston 2008). As we will show, the instructional format, in this case video, both makes available to analysis general aspects of following instructions and demands a form of work specific to the video medium.

In following instructions, there is a need to figure out what corresponds to what, to overcome the differences between what is shown, what is at hand, and what is to be done; to bridge the ideal world of instruction and the actual world we inevitably stumble over. In the context of practical, manual tasks, following instructions requires a form of understanding that necessarily goes through the physical experience of the described actions — a corporeal achievement which Livingston (2008:103) calls “embodied correspondence.” Inviting the reader to also follow the instructions and perform such activities as making origami or jigsaw puzzle, Livingston argues that any understanding of how a practical, manual task is performed requires the lived, embodied experience of it. As part of this lived experience is the physical experience of how objects react to and often resist our manipulations. Ultimately, experiencing embodied correspondence retrospectively provides the instructions with a meaning that, before engaging in action, could only be approximated.

Another central feature of instructions, such as manuals or recipes, is that tasks are decomposed into steps, actions or subactions to be accomplished one after the other. In textual instructions, this step-by-step decomposition is spatially organized and visible, so that moving from one step to another can be done by scanning through the text. Recent video-based studies of copresent instruction in contexts such as handicraft (Lindwall and Ekström 2012), sports (Evans and Lindwall 2020; Råman and Haddington 2018), driving (de Stefani and Gazin 2014; Deppermann 2019; Rauniomaa et al. 2018), cooking (Mondada 2014), and piloting (Melandar and Sahlström 2009), show that copresence makes it possible for instructors to design the instructions based on novices' actions. Instructors and novices' joint access to the situation and to each other's actions allows instructors to closely monitor novices' actions and thus intervene, specify an action, parse the task in smaller steps, correct mistakes, and so on (Lindwall, Lymer, and Greiffenhagen 2015). Users of video tutorials do not have an expert monitoring and adapting to their actions to adjust the

instructions to their performance, and therefore they need to find similar resources in the medium at hand.

Video Instructions

While instructional videos do not offer the same interactive resources as copresent instructional activities, they are artful demonstrations of tasks, finely coordinated descriptions of what to do and how to do it, combining verbal instructions with visual demonstrations. Relying on the many possibilities of video recording, such as shifting camera angles and zooming in and out, they provide access to the visible details of live actions, in the movement, and thus offer novices a lively view of the action. Moreover, like other instructional formats (Keevallik 2015), video instructions provide an initial parsing of the task, alongside the audio-visual and temporal unfolding of the video, made available to instruction followers through visible and audible cues.

As videos display the moment-to-moment progression of an instructional demonstration, their temporality does not necessarily match that of doing the task. Since the people following the instructions are typically not competent at the task, they are likely to take more time to perform each action than the instructor does in the video. In addition, the duration of the demonstrated task can be different from the duration it took the instructor to perform the task in the first place. In the videos that are available online, sections of the task are often cut off or accelerated through video editing in postproduction. Recent work has explored how video is produced (Ayaß and Gerhardt 2012; Broth, Mondada, and Laurier 2014). Some of this research looks at professionals at work, for example in live editing of sports shows on television (Engström, Oskar, and Mathias, 2010; Perry et al. 2019), reality TV parenting programs (McIlvenny 2011), or collaborative documentary editing by editor and director (Laurier and Brown 2014). Other studies deal with mundane video production; for example, Licoppe and Morel (2014) study how remote participants in video-communication move the webcam and articulate those motions with talk as they show their home interior. Studies of video production shed light on how producers expect and project the *reception* of their videos. This is also the case here — instructional videos are assembled for people to do a particular task in another place and time, which creates particular requirements. Producers therefore put their work and effort into designing videos that can support users' practical actions, for example by emphasizing the introduction of important objects or tools, or transitions between different phases of the task.

Clearly then, video instructions and their following involve multiple temporalities — the duration of actions in the video, the demonstrations' step-wise organization, and the time participants take to achieve each step — all of which need to be coordinated to create a consistent, progressing course of action. Those who follow the demonstrations have to make sense of the instructions while following them, and to match the progression of the video with their attempt to achieve the task that is being instructed.

Time and Temporalities

This makes the practical problem of managing time a starting point of this article, where we investigate some different ways of *pausing* video as the central resource. Measurement of time and temporal norms vary in history and regions of the world, which makes time a fundamental institution for social sciences. While it is a well-known fact that the hegemony of clock time originates in the industrial revolution, symbolic interactionism advocates that social times are complex and manifold (Adam 1995). Flaherty (2016), for instance, has argued that perception of time is a social and interactional achievement, and that “the willful modification of our own temporal experience is often realized through subtle and guarded practices” (Flaherty 2011:12). In a recent book review, Flaherty goes on to ask for “empirically grounded research that reveals how digital technology affects temporal experience in everyday life” (Flaherty 2016:505). These effects of technology on temporality have recently become a concern of work both on new media, and on the impacts of mobile technology (Gregg 2018; Julien 2019). What has mostly been lacking, however, and necessary to get to an understanding of time manipulation in situ, is a detailed approach of technological temporal practices, for example through the use of video media.

One important resource we draw on is discussions of the temporalities of talk in interaction. Conversation analysis as a field and as an analytic method, the substrate of this article, has paid close attention to this temporality, order, and timing of action. While early studies based on telephone data emphasized the sequentiality and progressivity of talk (Sacks, Schegloff, and Jefferson 1974), it has always been acknowledged that social interaction involves other temporal horizons and frameworks. Conversation analysis has addressed some of them in their own right, expanding from studies of talk in interaction to gaze and embodied conduct (Goodwin 1980; Heath 1986), objects and the material surrounding (Streeck, Goodwin, and LeBaron 2011), and practical tasks embedded in face-to-face interactions (Luff, Hindmarsh, and Heath 2000; Nevile 2007). Goodwin (2002:22), for instance, shows how participants in various settings are “attending to multiple temporal and sequential frameworks simultaneously,” and goes on to argue that “multiple temporalities embodied in diverse media are in fact a general, systematic feature of human action.”

Recent studies have focused on how different but interrelated temporalities coexist and might interfere with each other (Deppermann and Günthner 2015; Deppermann and Streeck 2018; Mondada 2018). In studies of multiactivity (Haddington et al. 2014a), the focus is on how “people organize multiple activities together, concurrently or serially” (Haddington et al. 2014b:5); for example, how hairdressers coordinate hair cutting and conversation with clients (de Stefani and Horlacher 2018), how office workers attend a phone call while talking with a copresent colleague (Licoppe and Tuncer 2014), or how volunteers coordinate massaging someone’s hand and conversing (Nishizaka and Sunaga 2015). These

studies show how embodied actions are coordinated with various interactional and material demands; how multiple activities are organized sequentially, serially, or simultaneously; and what interactional resources participants use in order to manage multiple tasks. As noted by Haddington et al. (2014b), there is a specific focus on practices of starting, restarting, resuming, continuing, on one hand, and stopping, halting, pausing, suspending, postponing, and abandoning, on the other. Given that the participants in our study organize watching the video and achieving the task as serial or parallel activities, and given that we focus on the practice of pausing, there are relevant similarities. However, as we come back to in the discussion, there are central differences and the phenomena therefore deserve separate treatments.

DATA AND METHOD

The article is based on analyses of video-recordings of ten dyads using instructional videos to achieve a variety of practical tasks. Because the use of instructional videos is often unplanned or responds to an emerging situation, this activity is difficult to capture in the form of naturally occurring data. A decision was therefore made to elicit data in a similar way as in Suchman's (2007) study of colleagues discovering how to use a photocopy machine through an expert help system, and Brooker and Sharrock's (2016) study of collaborative music making with a computer software. Observing several persons collaborating gives access to their practical reasoning as they verbalize their actions for each other.

One limitation of this approach concerns the occasioned nature of the recorded activities and the use of dyads, whereas instructional videos are most commonly used by persons doing the task on their own. One consequence of having dyads in action is that participants in the study tended to divide the task between them so that one of them would take charge of the video and the other would perform the task, a tendency perhaps exacerbated by their being asked to collaborate on the task. Nevertheless, this division of labor was flexible enough — participants would often swap between roles or jointly manipulate the laptop to navigate the video — for most if not all aspects of individual use to be available in the data.

We recruited ten pairs of participants through personal contacts and through a student volunteer website. All the participants gave their informed consent to being video-recorded and for anonymized transcripts of the video data to be used in academic presentations and publications. The participants were given a specific task with which they had little or no previous experience, and they were video-recorded for approximately an hour in a room equipped with three video cameras, an external microphone, and a computer with screen recording, as well as the tools and objects necessary to achieve the task. The data include the following tasks: replacing bicycle brakes; replacing a bicycle chain; picking a lock with paper clips; applying make-up with some unknown technique (two sessions); practicing yoga (two sessions); doing origami (two sessions); and cooking an unknown dish. All the pairs used a laptop

* ^ #	Delimit descriptions of speakers' actions, a specific symbol foreach participant
×	Placement of screen caption in talk or silence
bu-	Cut-off
(.)	A brief interval (about a tenth of a second)
(0.0)	Elapsed time by tenths of seconds
Grey	Description of bodily conduct

FIGURE 1. Transcript Conventions

and YouTube to search for and watch video tutorials of their choosing. While only some of the participants are native English speakers, all the data is in English.

We annotated the video recordings with the Elan video annotation software, to locate when a video is playing, when it is paused, and when participants are scrubbing (i.e., moving the play button in the timeline). This initial annotation, a first step in the analytic process, gave us an overview of participants' interactions with the video — when they paused, moved the timeline, and pressed play. Intrigued by how often participants paused the video, we looked at the 150 instances of pausing. We eliminated cases where they replayed a section of the video after they had either failed to understand the instruction or encountered a problem while attempting to do what the video instructed, the type of situations investigated by Heinemann and Möller (2015). The latter is (to the best of our knowledge) the only similar, video-based study of instructional videos in use. Studying novice knitters, the article shows how they locate the source of past mistakes by moving backwards in the video, and navigate the video to start again where they did wrong. We decided to focus on moments where participants were playing a section of the video for the first time, before or while attempting to achieve the demonstrated actions themselves. In those circumstances, pausing appeared to be done mainly to coordinate the temporality of the video with that of the practical task.

The instances presented in this article have been transcribed following Jeffersonian conventions for talk (Jefferson 2004) and Mondada's conventions (Mondada 2018) for embodied conduct, in addition to which a horizontal line indicates when the video is paused or resumed (Figure 1).

The Phenomenon

We systematically analyzed a collection of fifty two instances of participants pausing the video to coordinate the temporality of the video and that of the physical task.

Three distinct types of pausing emerged: (1) Pausing to address a correspondence problem: participants paused the video instructions to raise and address potential discrepancies between their material environments or objects and those of the video instructions; (2) Pausing to keep up: when participants accomplished their task while the video was playing, they needed at times to pause the video instructions to catch up with the demonstration; and (3) Pausing to turn to action: when participants alternated between attending the video and performing the task, they paused during transitions between steps in the video in order to reproduce the demonstrated step.

This typology is not meant as a members' object; however, it proves robust as a way of ordering the materials, and it is overall representative of the data. Indeed, each type of pausing has specific characteristics and all of the fifty two instances of pausing correspond to one type. The aim of this typology is to shed light on the coordination of temporalities — why then and what for. The empirical section is accordingly organized in three sections, with two excerpts analyzed in each section.

PAUSE AND MAKING TIME: EMPIRICAL SECTION

Pausing to Address a Correspondence Problem

Instructional videos recurrently introduce or refer to objects, implements, different parts of an object, and so on, before or while demonstrating actions with those objects. As discussed at length by both Suchman (2007) and Garfinkel (2002), however, in actual cases of following instructions there can be significant and potentially problematic differences between the world of the instructions and the world of those who follow them. Before the participants in our study attempt to follow the demonstrated actions, they often struggle to match the objects used in the video with the ones at hand. What they see and hear descriptions of in the video does not always immediately correspond to what they have at hand. Participants might consider that their objects are different in nature, and therefore need time to work out how to overcome these differences and follow the instructions anyway. Even in cases where they consider their object as identical or similar, they might still have difficulties identifying specific aspects of the object referred to in the video, or how to use the object.

Borrowing from — and slightly bending — Livingston's notion of "embodied correspondence," we call this the "correspondence problem." When a correspondence problem arises, participants regularly pause the video to make time to solve it. By momentarily suspending the video, they adjust the progression of the instructions to the progression of their task. Before proceeding with the video, they then attempt to establish a correspondence between what they just saw in the video and what they have at hand. This type of pausing occurs in the course of an in-progress instruction, and participants orient to and account for this interruptive character through talk. The correspondence problem is formulated and specified, through talk and embodied actions, such as pointing gestures toward objects or elements in the video and/or in the physical world, which creates a comparison or contrast between the video and the physical space.

Excerpt 1 involves Ali and Molly replacing the brakes on a bicycle. In reading the transcript, it is important to distinguish the instructor's voice recorded in the video and played through the computer speakers (INS) from the talk of the two participants (MOL and ALI). Previously, Ali and Molly replaced the brakes on the front wheel, after which the instructions moved on to the next step: how to balance the spring tension on both sides of the calipers. As the excerpt starts, Ali is sitting on one side of the bicycle wheel, Molly is standing on the other (Image 1.1a), and they are attending the video where the instructor begins to unpack this step.

The instructor specifies how to balance the spring tension with *and the way that we do that, is by adjusting (.) the adjusting screw (.) on (.) this spring tension* (lines 1–2) while pointing to a specific part of the caliper, allegedly the screw (Image 1.1b). During the last part of this utterance and while the instructor maintains the pointing gesture, Molly bends toward the computer, and pauses the video (Images 1.2 a and b) while the instructor initiates the next utterance (line 3). Molly immediately turns to the bicycle, and with the interruption marker (Schegloff 1987) *so wait* (line 5), displays that she is interrupting an instruction in progress. Then, with *that one is this* (line 5) pointing to a screw on the caliper, she raises a correspondence problem and asks her coparticipant whether he has the same understanding about which screw the instructor is referring to, on their bicycle. Thus, by pausing the video at this point in time, she makes time to collaboratively investigate the problem, involving Ali.

Ali finger points to another screw on the caliper and contradicts Molly with a different view: *it's this thing* (line 7, Image 1.3). They continue to disagree (lines 9–10), and on line 10, Ali suggests that Molly plays the video again to establish correspondence simply by seeing what comes next in the video (by realizing she is wrong, which, eventually, she turns out not to be): *well (.) you will see him do it just watch*. Shortly after, Molly resumes the video and they establish correspondence (not reproduced), possibly in part because the change of camera viewpoint allows them to better distinguish between the different screws (Figure 2).

In the excerpt, we can see how the participants rely on the video, in particular the instructor's embodied actions and verbal intonation, but also the camera viewpoints. The marked shape of the instructor's pointing to the screw conveys precision and invites participants to identify one screw in particular. The way Molly pauses at the end of the instructor's *this spring tension* (line 2) responds to this indication, as well as to the instructor's falling intonation, which indicates the completion of this step and projects transition to the next step. Current speakers' falling intonation coupled with grammatical completion are commonplace cues for coparticipants in interaction to anticipate transition spaces for turn-taking (Sacks, Schegloff, and Jefferson 1974). Similarly, the instructor's use of intonation in the video enable instruction followers to anticipate the completion of an action or series of related actions, and transition to the next one. Relying on such cues, Molly pauses at this moment in order to make time to establish correspondence. Ali shows a different analysis of how the instructions are organized and unfold: the way he suggests to proceed with the video in order to find out about the screw shows that he expects more to come about this

01 INS and the way that we do that, is by adjusting (.)

02 INS the adjusting screw* (.) on (.) this spring *tension.

ima *Images 1.1 a and b

mol *bends to computer-->

03 INS the s-*

ima *Images 1.2 a and b

mol *pauses video, turns to bicycle-->

04 (1.2)

05 MOL so wait.* (.) *that one i:s *this

mol -->* *-----*points to side crew, taps-->

06 ^ (1.0)

ali ^moves left arm to brake-->

07 ALI it's ^this thing*

ali -->^points to other screw

ima *Image 1.3

08 (1.9)*(0.5)^(0.5)

mol *turns head to video

ali ^withdraws pointing, turns head away

09 MOL ne: it's this ^thing

ali ^turns head to video

10 ALI *no (.) well (.) you will see him do it just watch*

mol *turns head to brake, inspects *bends closer

EXCERPT 1.



FIGURE 2. Different Viewpoint in the Video Shortly after Images 1b and 2b

referent before the instructions move on too far, so that correspondence will be easy to establish by playing the video a little further. This prospective orientation is integral to following instructions broadly: we tend to look for what comes next to make sense of what we are doing now, as a part of the task's broader gestalt.

In Excerpt 1, the object participants have at hand is very similar to the one in the video, and the correspondence problem is a matter of distinguishing between two small elements. Excerpt 2 shows a different sort of correspondence problem. Emma and Clara do not have the same products as the ones used in the eye-brow make-up tutorial, and they therefore need to find a similar object among those at hand which they could use instead. The analysis focuses on Emma's actions.

As Excerpt 2 starts, Emma has finished brushing her brows upwards and she is fully attending the video (Image 2.1a). The instructor initiates a new step by introducing a new product: *so: I like to use my: anastasia beverly hills dip bro-* (line 1). Shortly after, the instructor brings the pomade in the frame (Image 2.1b), Emma pauses the video, points to the product on the screen (Image 2.2), and identifies it with *okay so she has (.) the p- pomade (.) thing* (line 2). She apparently knows from the outset that they do not have the same product, and raises an objectual correspondence problem — instead of referential as in Excerpt 1 — asking if any of the products they have at hand can be used instead of this pomade. The token “okay,” pervasive in our data, both claims understanding and is transitionally relevant, responsive to prior talk while simultaneously setting up next-positioned matters, acting as a transition marker (Beach 1993). By emphasizing she and turning her head to the table where their products are while maintaining her pointing gesture to the screen, Emma builds a comparison and raises a discrepancy between objects available in the video and objects available in their world.

The way the instructor verbally introduces the product she is going to use next, and the way she brings it into the camera frame to make it clearly visible, can be treated by participants as an indication that they should get a hold of it now. Emma relies on these cues in the video to pause at this moment in order to raise the correspondence problem, and make time to find an equivalent product.

After closing the inserted sequence initiated on line 3, Emma repeats her trouble-raising turn from line 2 with the same pointing gesture (Image 2.3), and

01 INS so: I like to use my: #*anastasia beverly *hills *dip bro-
 ins #brings product in video frame
 ima *Images 2.1a and b

emm *reaches to touchpad
 *pauses video

02 EMM *okay so she has * (.) * the p- pomade* (.) thing.
 emm *points to screen*turns head to table*withdraws pointing
 ima *Image 2.2

03 CLA did she just brush them upwards like this
 04 EMM hh I dhon't kno:w hh
 05 CLA hh hhh hhh
 06 EMM that looks wei- yea:h she di:d
 07 CLA okay nice.
 08 EMM nice? Okay sh-
 09 EMM so *she's* using the pomade. *should we?*

emm *points to screen *withdraws pointing, turns head to table
 ima *Image 2.3 *Image 2.4

10 (1.8)
 11 EMM should we still do this?
 12 EMM I mean I think *everybody's gonna use this *product=
 emm *points to screen. *withdraws pointing
 13 EMM because it's insta [(.) friendly]
 14 CLA [yeah let's try] with [what] we've got
 15 EMM *[we]
 emm *points to product-->
 16 EMM we could *do this* (.) *I think.^
 emm *turns head to screen
 *withdraws pointing
 cla ^plays video

17 EMM [it's basically the same.]
 18 INS [and a medium brojwn, (.) and eye[brow brush that I] always use
 19 CLA [let's see what she does.]

EXCERPT 2.

turns to their products (Image 2.4). *Sh- so she's using the pomade* (line 9) is now preliminary to the unspecific, open-ended proposal *should we? (1.8) should we still do this? I mean I think everybody's gonna use this product because it's insta friendly* (lines 9–13). Partly on overlap, Clara and Emma both propose to continue, Clara relying on the feasibility to find something usable with *yeah let us try with what we have got* (line 14), and Emma with an incrementally more specific proposal: *we could do this (.) I think* (line 16), pointing to a box of powder on their left (Image 2.5), and *it's basically the same* (line 17). Clara plays the video again (line 16), and her proposal *let us see what she does* (line 19) shows a similar orientation as in Excerpt 1 to finding out what to do by looking to what comes next in the instructions.

To sum up, pausing the video can be a resource to coordinate the temporality of the task with that of the video by making time to compare and establish correspondence between objects in the world of the instructions and objects in the physical world. By pausing the video, one participant orients to the matching of objects as an issue that needs to be resolved before the video goes any further. The pause is not anticipated or prepared, it occurs outside transition spaces, while an instruction is in progress in the video, after and in response to the video introducing a new object or referring to some material feature. Participants typically point to objects onscreen and in their environment, thereby referring to objects deictically and indexically through environmentally coupled gestures (Goodwin 2007). These embodied, verbal, and contextual configurations build and make salient a comparison between the what is shown in the video and what they have access to in their surroundings.

Parallel and Alternate Organizations

Before we move on to the analyses of the next two types of pausing, it is necessary to introduce two different ways participants organize the watching and following of video instructions. In what we call the *parallel* organization, participants attempt to achieve the task while the video is playing, and they coordinate the temporalities by pausing the video to keep up with how the task progresses. In contrast, in the *alternate* organization, participants alternate between attending to the video only, pausing, then attempting to achieve the task while the video is on pause. In this case, they coordinate temporalities by pausing and turning to action, round about when the video has completed an instruction for one step, before it moves on to the next step. Pausing to keep up with the video and pausing to turn to action are thus specific to parallel and alternate organizations respectively (while pausing to address a correspondence problem can occur in both organizations).

Some activities allow for the parallel organization simply because participants keep their hands close enough to the device where they can pause. Other activities preclude it, such as doing yoga, because participants are either too far from the device, and/or they are physically engaged in performances away from the device. With the exception of the physical arrangement and constraints integral to the task, the relation between the task and the approach is not straightforward. In our data,

the same dyad might take both approaches in the course of their task, and different dyads performing the same task could take either approach. This was particularly the case for picking up locks and applying make-up, whereas the alternate organization was never taken in doing yoga.

Pausing to Keep Up

Participants can achieve the task while the video plays — that is, follow the video instructions in both the sense of understanding and applying them practically as they are delivered. In this parallel organization of the activity, participants' gaze moves back and forth between the video and their manual actions. Although the participants' attempts to achieve the tasks typically unfolds with a stepwise organization similar to the tasks in the video, their actions often do not have the same duration as those in the video. When the video continues with a new step before the participants have completed the previous, it can be hard for them to keep up with the instructions. Pausing the video thereby becomes a way to coordinate the instructions to the development of the task at hand.

Excerpt 3 shows Ali and Molly earlier in their task: they are installing the new brake pads, each kneeling on one side of the front wheel (Image 3.1a), meanwhile that very same step is being demonstrated in the video. On lines 1 to 4, the instructor formulates and demonstrates (Image 3.1b) three consecutive actions part of that step: installing the spacers, tightening the bolts, and aligning the pad with the wheel curvature (lines 1 to 3). Molly and Ali are currently tightening the bolt, thus close to completing the step currently demonstrated; they are also looking at what they are doing, not at the video.

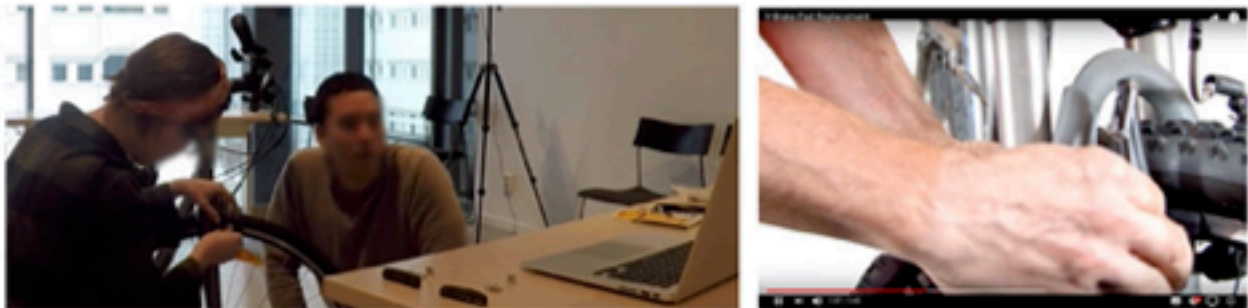
While the instructor verbalizes *and just very approximately (.) line it up* (line 3), Ali is visibly completing this step: he removes his hands from the brake to his laps, and turns his head to the video (line 3, Image 3.2a). Molly remains fully engaged with her manual actions. Shortly after, the video zooms in on a new tool (Image 3.3) and the tutorial moves to the next step. As the instructor names the tool (*this little brake shoe tuner* line 5), Ali bends over to pause the video (Image 3.4). He turns his upper body to the bicycle again, marks a transition with *°okay°* (line 7), and bends over the wheel to look at how Molly is doing with inserting the pad (Image 3.5). In other words, he shows that he is done with installing the new pad and waits for Molly to also be done. By pausing at this moment, he is making time for both of them to complete this step, in order not to fall behind the video as the instructions are moving on to the next step.

A series of typical cues in the video displayed this transition to the next step: the instructor's expanded *and* and falling intonation on line 3, the change of camera viewpoint shortly after, the showing of a new tool and its verbal introduction (line 5) with a noticeable change of tone in the instructor's voice. Ali pauses as soon as possible after the video has moved on to the next step, while an utterance is still in progress in the video. He makes a rapid motion to the space bar, in response to the

01 INS insta:ll all the spacers, (1.2) and just t- hand (.)
02 INS tighten (0.5) initially.* (1.4)
ima *Image 3.1 a and b



03 INS a:nd just *very* approximately* (.) line it up.
ali *-----*turns to video-->
ima *Images 3.2 a and b



04 (0.2) # (0.8)
vid #image changes
05 INS *this little *brake shoe tuner* (.)* is a-
ali *bends over *pauses video
ima *Image 3.3 *Image 3.4



06 * (0.5) * (0.2)
ali *-----*withdraws arm, turns to bicycle
07 ALI °okay°*
ali *bends closer to Molly's actions
ima *Image 3.5



EXCERPT 3.

initiation of the next step in the video, and in relation to Molly's ongoing involvement in the previous step. This type of pausing embodies a form of coordinating temporalities where delivery of video instructions is suspended every now and then to catch with the progression of the task in the video.

In Excerpt 4, Emma and Clara have moved on with the make-up tutorial, and our analysis again focuses on Emma. She alternates between looking at her actions and the screen while continuing her task, letting the video move on to the next step while she both actively attends the instruction and progresses the task from the previous step.

In this excerpt, Emma pauses a bit further into the next step than Ali does in Excerpt 3, and as she pauses, she orients to these different steps of the task as equally relevant: one that she is going to complete soon, the other she is already anticipating and getting ready for. In other words, she manages a larger delay (compared to Excerpt 3) between the progression of the video demonstration and that of her task. As the excerpt starts, Emma is applying some powder from the tip to the inside of her brow with a small brush, looking at her actions in the mirror (Image 4.1a).

While the instructor expands on the step Emma is currently accomplishing (*and I like bringing it in a little bit further than you think you would, because* (0.4) lines 1–2), Emma suspends her brushing actions to turn her gaze to the video (Images 4.2 a and b) and then turns her gaze back to the mirror before *once you do that* (line 2). When the instructor verbally introduces a new tool (*>I now take my foundation brush where I applied my foundation with* < line 3), Emma suspends her brushing actions again and shifts her gaze to the video. After this, the instructor shows the big brush, verbalizes the action to be done with it, and taps the front of her brow with it (*and then just lightly with that foundation brush just tap the front of the brow*, lines 4, 5, and 9, Image 4.3).

As shown by her tokens of appreciation (lines 6–8), Emma is actively following the instructions. By shifting her gaze between the mirror and the video while brushing her brow, she is able to follow the instructions for the step coming after the one she is currently doing. While doing this, she also orients to the upcoming step by visually searching for a big brush similar to the instructor's on her left (line 8). She then brings her left hand to the keyboard (Image 4.4a), turns her gaze to the video again, and pauses while the instructor expands on this action with the big brush (*and that will softe-* line 10). She turns to the mirror again to continue her brushing action once the video is on pause (Image 4.5), while planning for the next step by proposing a tool they could use in place of the big brush in the video (line 12).

As in all instances of pausing as keeping up, Emma makes time to complete the ongoing step after the video has moved on to the next. She also interweaves watching the video instructions and achieving the task by shifting her gaze and suspending her actions when she attends to the video; and she interweaves several steps of the task. Instead of pausing the video right after the initiation of a new step, she actively attends to what comes next for a few seconds, so that she can plan what she is going to do after.

01

INS


And I like bringing it* in a little bit *further than

ima

*Images 4.1 a and b

emm

*turns gaze to video



02

INS

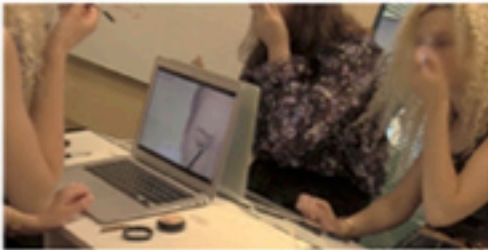
you think* you would, because (0.4) *once you do that,


ima

*Images 4.2 a and b

emm

*turns gaze to mirror, taps brush





03

INS

>I now take my foundation brush where *I applied my foundation with<

emm

*turns gaze to video-->

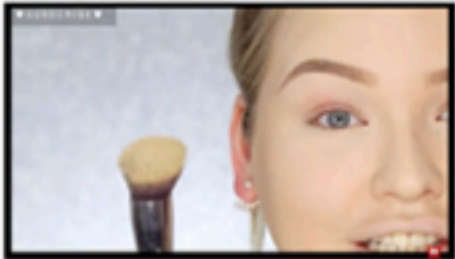
04

INS

and then just* lightly with that foundation brush

ima

*Image 4.3



05

INS

#just t[ap the [front]

ins

#taps

06

EMM

[uhu:n

07

CLA

[mm:]

08

EMM

mm[m: *that looks]* good

emm

*turns gaze to left, brings LH to keyboard


09

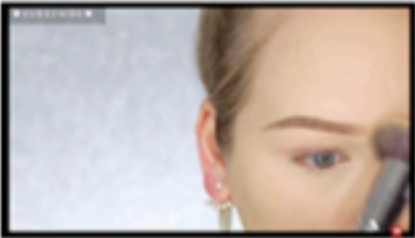
INS

[of the brow,]

ima

*Images 4.4 a and b





10

INS

>and *[that will<] *softe-

11

CLA

[yeah]

emm

*turns gaze to video

*pauses video

12

EMM


**okay* (.) so: (.) we could use (.) just (.)

emm

*turns gaze to mirror, resumes tapping actions

ima

*Image 4.5



13

EMM

the concealer again because it was a good color

EXCERPT 4.

Although she pauses after the completion in the video of her ongoing step, and further into the next step than in Excerpt 3, Emma can be seen to rely on at least one similar cue in the instructor's talk indicating a form of completion: *and then* (line 4) framing the following action (*just lightly with that foundation brush just tap the front of the brow*) as final. But she also disregards the slightly rising intonation on *brow*, (lines 4, 5, and 9) which projects more to come, and this probably relates precisely to her already falling a little behind and needing to pause before she gets too far behind.

In this section, we have analyzed cases where participants proceed with their embodied and material task while the video played in parallel. In order to adjust the progression of the video to that of their task, they temporarily suspend their flow to catch up with them. In the next section, we analyze cases where participants do not engage in their physical task while the video plays: either they watch and listen to the video, or they achieve the physical task while the video is on pause. We show that in this alternate organization, participants pause in order to turn to their practical task when the instruction of a step is complete. Their pausing thus displays their understanding of the parsing of the task in the video and of the task's internal organization; it also achieves an additional parsing of the task.

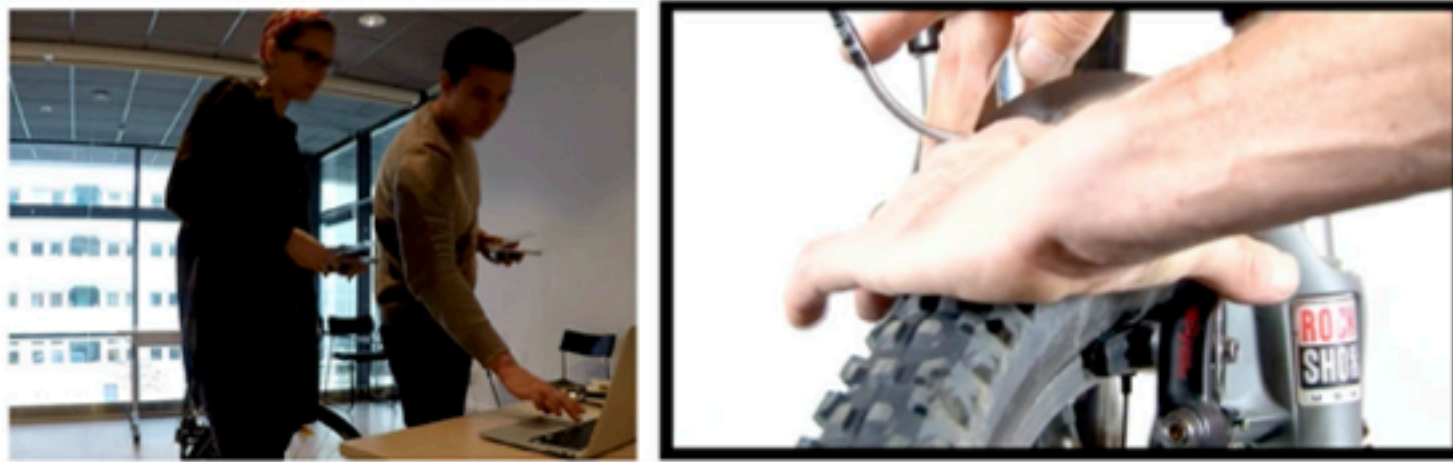
Pausing to Turn to Action

One way of attending to the task is to alternate between either watching the demonstration or doing the task while the video is paused. When participants use the video instructions in this way, they coordinate the temporality of the instructions to that of their task by segmenting the video into practically relevant, reproducible steps. The participants pause when they have enough instructions to fully achieve one step of the task, typically at the end of a step in the demonstration and before the next one begins.

As shown in the previous sections, the video instructions provide an initial parsing of the task into steps. Reflecting this, there are a number of cues or indications embedded in the video instructions on which the participants rely to pause, such as the instructors' phrasal intonation. In pausing to turn to action, participants rely on these cues to target transition spaces to pause. Parsing the task in the video to turn to action also requires an understanding of the task's internal organization: some parts can and need to be separated, whereas others cannot. This type of pausing, thus, exhibits participants' understanding of the video instructions as a series of both separate and interrelated parts.

Excerpt 5 takes place before Excerpt 2, back with Molly and Ali. As the excerpt starts, they are both fully turned to and attending the video with the bicycle behind them. The analysis focuses on Ali who both interacts with the video device and takes the lead to organize the activity. On line 1, the instructor announces the task that will be instructed (*let us replace the pads on these brakes*), which he subsequently unpacks into smaller steps.


```
01  INS      let's replace the pads on these brakes. (0.9)
02  INS      in order to do that, (1.0) we need to (.) release,
03  INS      the quick release mechanism *on the cable,*x
ali                                     *-----*bends, RH to keyboard-->
ima                                     *Images 5.1 a and b
```



```

04  INS      which is done by opening up the little rubber boot,
05  INS      a:nd (.) pulling (0.5) the calipers together, (0.8)
06  INS      a:nd unhookx (1.5) this xL-shaped (0.7) tube, (1.0)
      ima      xImage 5.2  xImage 5.3

```



```
07  INS      called the noodle.* (.) it-
    ali      *pauses video, turns to bicycle-->
```

```
08  ALI      okay× (.) let's do that
    ima      ×Images 5.4 a and b
```



At the end of the instructor's grammatically complete utterance *we need to (.) release the quick release mechanism on the cable* (lines 2–3), and as the action in the video looks almost complete (Image 5.1b), Ali bends toward the computer and gets ready to pause by placing a finger above the space bar (Image 5.1a). He stays in this unstable, expecting position while the instructor unpacks again the step from lines 2–3 into a series of smaller steps: *which is done by opening up the little rubber boot, a:nd (.) pulling (0.5) the calipers together (0.8) a:nd unhook (1.5) this L-shaped (0.7) tube* (lines 4–6). After the instructor utters *unhook*, the unhooking is visually demonstrated (line 6, Image 5.2), and on *L-shaped*, Ali moves his finger up and down above the touchpad (Image 5.3). First, placing his finger above the space bar, and then this movement up and down above it as an aborted attempt to pause, show that Ali expects the completion of this step to be imminent, which would allow them to turn to action and achieve those same actions.

Through pauses and elongations in his talk, the instructor adjusts verbal instructions to manual demonstrations, extending the former to match the duration of the latter, one form of coordination of temporalities in the instructions themselves (Keevallik 2015). The instructor ends each short utterance on a rising intonation (for example *a:nd (.) pulling (0.5) the calipers together*, line 5), which projects more to come. Ali demonstrably relies on this rising intonation to not pause yet, overlooking the syntactical completion projected by several, subsequent *a:nd* (lines 5 and 6). The way he does not pause yet is also sensitive to the internal organization of the task. The demonstrations make visible that between “pulling the calipers together” and “unhook[ing] this L-shaped tube,” one needs to maintain the pressure with thumb and index since the calipers are by default mechanically pulled apart by a spring. It would not make sense to pause and turn to action between two inseparable actions.

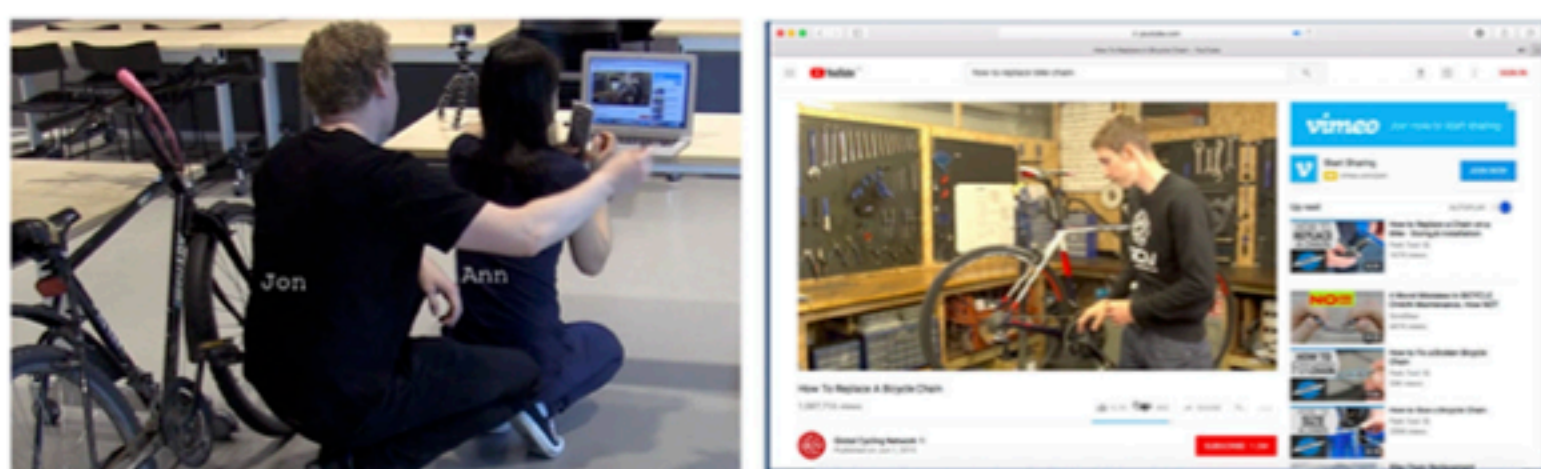
Ali pauses exactly at the end of the instructor's utterance *called the noodle* (line 7), which falling intonation stands in stark contrast with the former rising ones. He immediately turns away from the computer, to the bicycle (Image 5.4), and as his proposal *okay (.) let us do that* (line 8) makes explicit, they are now turning to action to do what the video just demonstrated.

Throughout the excerpt, Ali relies on the instructor's intonation and on his understanding of the task's organization. He makes use of the instructor's unpacking of the task into smaller steps, and pauses to turn to action in a transition space between the first two steps which can be separated. But while targeting transition spaces is a general orientation when pausing to turn to action, participants do not always pause exactly in those spaces. In Excerpt 6, Ann and Jon have not yet started with their task, replacing the chain on a bicycle. They are watching an instructional video beforehand, crouching in front of a table where the computer is set, with Ann closer to the computer and Jon to the bicycle behind them. At the beginning of the video and shortly before the excerpt starts, the instructor introduced the task and the objects it requires. As the excerpt starts, Ann has just gotten hold of the new chain, and the instructor introduces what is hearable as the first step in the overall task with *To remove the old chain* (line 1), and thereby projects what is to come as a series of steps.


```

01  INS      to remove the old chain, *firstly start by putting the gears
      ann          *takes chain tool from table
02  INS      in the ^smallest cog at the back,*
      jon          ^takes chain tool from Ann's hand
      ima          *Images 6.1 a and b

```



```

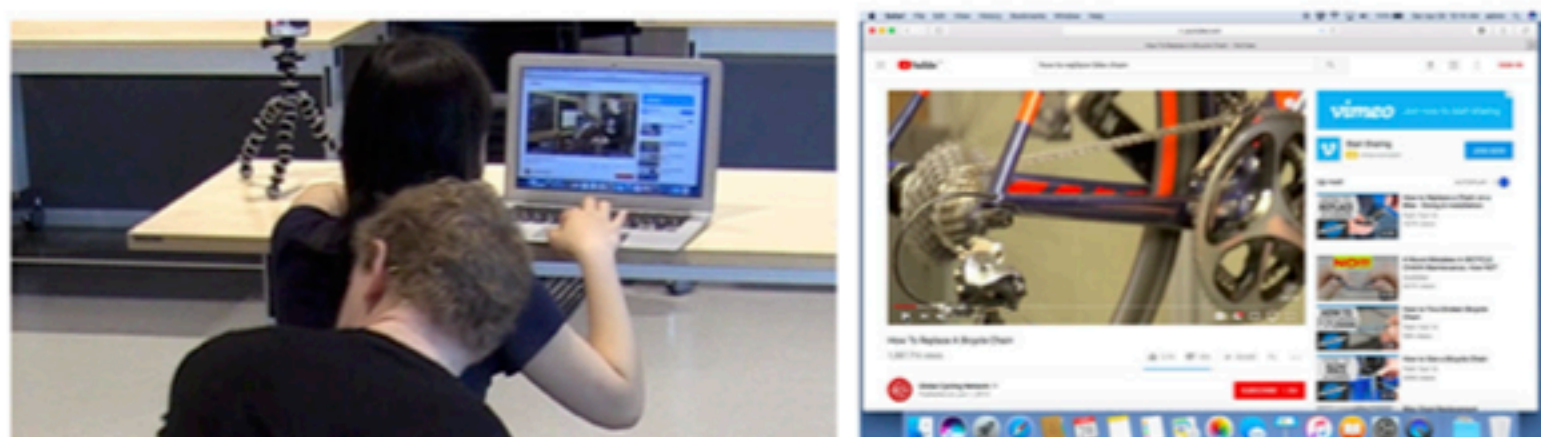
03  INS      *and also the smallest chainring in the front.=
      ann      *hand to touchpad-->

04  INS      ^=this *xgives you the most *possible amount of s-x
      jon      ^turns away from screen to bicycle-->

      ann      *moves cursor          *pauses video

      ima      *Image 6.2                                     *I

```



05 ANN right. *so:* we can put it on the smallest
Ann *turns away from screen to bicycle-->

EXCERPT 6.

The instructor decomposes *to remove the old chain* (line 1), the first step of the overall task, in smaller steps with *firstly start by putting the gears in the smallest cog at the back, and also the small chainring in the front* (lines 1 to 3, Image 6.1b), thus incrementally increasing the granularity. Meanwhile, Jon takes the chain tool from Ann's hand (line 2, Image 6.1a), and shortly after Ann moves her hand to the touchpad. Like Ali in Excerpt 5, she is getting ready to pause the video, thus displaying her anticipation of an upcoming completion and transition space.

As the instructor completes *and also the small chainring in the front* (line 3) on a falling intonation, Jon turns away from the screen (Image 6.2). He is already turning to action, relying on the same projective cues as Ann, and probably relying on her to

pause the video a distribution of work which is implied from the start through their relative body positions.

Ann continues to move the cursor toward the pause button, and pauses shortly after the completion point in the video, during the instructor's next turn hearable as an expansion providing additional background to the former action (*this gives you the most possible amount of s-*, line 4). With *right. so:* (line 7), Ann acknowledges the video demonstration and marks the transition to action while turning away from the computer to the bicycle (Image 6.4). She then rephrases the instruction from line 2 with the proposal (Couper-Kuhlen 2014) *we can put it on the smallest* (line 7), thus organizing their task about to start now.

In a slightly different way than in Excerpt 5, where Ali paused exactly on the instructor's falling intonation and within the transition space, the pausing here occurs when a new utterance is already underway. Through her next turn-at-talk, however, Ann parses the task in a way that aligns with Jon's turning to action, and disregards the latter section of the video. Most participants in our data use the space button to immediately pause the video. This is one of the few instances where pausing is done by moving the cursor to the pause button. Pausing by moving the cursor and clicking takes considerably longer than with the space bar, especially with the computer touchpad (unlike with, say, a mouse). As a consequence, Ann's pausing could be seen as delayed by the way she physically interacts with the device: Ann and Jon's parsing of the task is exhibited not so much in the exact moment they pause as in when Ann prepares to pause, when Jon prepares to turn to action, and how they initiate their actions, by organizing it prospectively.

In this section, we have shown how participants who alternate between attending to the video instructions and achieving the task coordinate the temporalities of the video and the task by pausing in or shortly after transition spaces. Participants wait for, expect, and anticipate when the video instructions reach a completion point, completing an action or series of actions that will be reproducible at once. They aim to pause before instructions for the next actions begin. In the data presented here, instructors' intonation as well as grammatical constructions are central resources for the ways in which the participants pause the video and parse the tasks. They also rely on their ongoing, moment-by-moment discovery and understanding of the task's internal organization. In cases where they do not pause exactly within a transition space in the video like in Excerpt 6, they can display their own parsing of the task as they turn to action. That they pause at the next possible opportunity suggests an orientation to preserving the gestalt of each step of the task and the structural relation between them.

DISCUSSION

This study has focused on the making of time and coordination of multiple temporalities. Bergson (2001) famously pointed to subjective experience of time as incommensurable with measured, normative times. A more sociological way of approaching

time is to look at *when* and *how* it becomes an observable concern in activities and interactions. Coming back to Flaherty's (2011:2) point from the introduction, there are plenty of "subtle and guarded practices" available out there for study where people can be seen to make time and shape their perception and experience of it. Moreover, interactive media and mobile technologies in particular bring in temporalities of their own, especially through how video flows, and our devices' constant connectivity to these media sources. By getting to grips with video-recorded instances of video in use, the type of study conducted here gives us a way into people's ordinary methods used to domesticate those devices and media, to embed the latter's partly autonomous temporality in the colocated, lived temporality, while experiencing the material world through one's body and with others. While these technologies affect our experiences, we act upon them in turn, and through concrete manipulations we participate in how they shape our time.

Our study teases out the complex relationships between what is done with the device *and* around it altogether, with a physical task proper as a central concern. Its import spans beyond this case study since threading together computer interfaces, the temporal unfolding of what they display, and a task being carried out in the colocated space, is pervasive in copresent and technology-mediated activities (Luff and Heath 2002; Szymanski 1999), where actions of suspending, making time, or resuming are also recurrent. In our digitally mediated lives, we commonly juggle between devices and the surrounding world without even noticing, when we use smartphones and apps, from webpages held up in discussions (Brown et al. 2018) to digital maps used in spatial navigation (Brown and Laurier 2005). By weaving together watching the onscreen instructions and achieving their practical task, our participants make one coherent and seamless course of action out of the two. Watching and doing are inseparable: they support each other, they do not compete with one another. This form of involvement and what we want to emphasize with this study stand in stark contrast with the competition and incompatibility between involvements emphasized in research on multiactivity and multitasking. Whether this contrast is integral to the activities under study or a by-product of the research focus is an open question.

Online video instructions and their following can also be seen as a collaborative activity of a special sort, starting when people who produce these videos video-record a task while it is being performed and demonstrated. Then, as they edit the video, they transform a "raw" version of the task in a more or less elaborate instructional video, before uploading it onto well-known platforms where users will easily find them. In another time and space, thus, a user will select this video, trusting its pedagogical qualities to be able to perform the task themselves. Although beyond the scope of this article, we found in our data how video producers creatively relied on a variety of camera edits and inserted text to draw users' attention to transitions between different steps of the task, and thus facilitate their work of coordinating video and task. In another publication (Tuncer, Brown, and Lindwall 2020), we explore how this form of mediated collaboration could also be sustained by video tools and players. In brief, the activity we focused on in this study as locally and temporally bounded

can also be seen as one site of an expanded chain of actions, where the final performance relies on “a network of people who cooperate to produce that work” (Becker 1974:767), thus echoing Becker’s conception of art as a form collective action, where the final work results from different professionals doing their bit in different times and different places.

CONCLUSION

With this video-based study, we aimed to provide an understanding of how such a commonplace and yet under-researched everyday-life practice as following an instructional video is concretely done. We were able to show in actual instances how participants coordinate multiple temporalities, mainly that of the video instructions and that of their practical task, which bring together dimensions of duration and sequential progression. While the participants’ main resource for this is to pause the video, close inspection of the actions around pausing shows that it can embody different concerns and orientations to the task underway and its relation to the instructions.

First, we showed how pausing could be used to address problems of correspondence between the task and the video. While an instruction is in progress, and often immediately after a new object has been introduced, participants can pause the video in order to establish correspondence between objects in the video and those they have at hand. After pausing, they account for halting an in-progress instruction by stating the correspondence problem and showing that they are making time to try to resolve it. Second, we showed that when participants achieve the task while the video plays, they can pause the video before it gets too far ahead of their action, in order to make time to catch up with the instructions and keep up with their progression. This is typically occasioned by, and takes place after, the demonstrated completion of one step and the move into the next one. Third, when participants alternate between attending the video and achieving their task, they pause the video once the instruction for one step of the task is recognizably complete. As the participants tend to pause during transition spaces between two steps of the task, this type of pausing makes visible participants’ online analysis of the instructions and internal organization of the task and how they parse the task into manageable segments.

In relation to this, we also showed how coordinating temporalities relies on cues in the video. In pausing to raise a correspondence problem, participants understand that a new tool or object is being referred to through instructors’ talk, and often with the object’s visibility emphasized on the image, and they pause to make time and get hold of or identify such objects. In pausing to keep up with the video, participants recurrently rely on cues indicating the start of a new step, since they aim to pause before the video goes too far ahead, in order to complete the current step. In this type of pausing as well as in pausing to turn to action, though, central cues include the instructors’ intonation and the emergent grammatical construction of their utterances which project and indicate transitions between different steps of the task. An

alternate task organization requires careful attention to when and where transitions are made, and when it will be necessary to pause the video to move to the task. In contrast, a simultaneous organization will demand a user to manage both the video and task at the same time, drawing particularly on the audio commentary to the video, but also on the ability to physically arrange the task such that moving between the two will not be physically too difficult.

Our study, as the title also aims to reflect, nicely illustrates Garfinkel's (2002) distinction between marking and making time, based on a musical experiment. As one plays a piano piece, one makes time — for the piece, while a metronome can constrain the time one makes by marking it, so that “With the pair we are in the midst of an organizational thing: we can't take all the time in the world to play the prelude” (Garfinkel 2002:99). While instructional videos and their autonomous temporality tend to mark time and impose a pace on the realization of the task, by pausing the video participants can make time to achieve different types of works involved in doing the task.

Lastly, this study is an attempt to combine interactional analysis with a focus on practical reasoning. While ethnomethodological studies investigate following instructions using ethnography, in a first-person perspective, or using re-enactments, and while video-based research on instructional settings focus on interactional accomplishments, we proceeded with the latter's systematic, analytic approach to investigate practical reasoning in following instructions, rather than collaborative interactions. We analyzed participants' interactions with each other and with the video device, an approach more common in human-computer interactions (Suchman 2007) and workplace studies (Heath and Luff 2000) aiming to shed lights on practical reasoning through the natural accountability resulting from the collaborative situation. We intend this article also as a call for more studies of this type which analyze interactional moves as accounts for actions, and thus investigate action formation and practices beyond interactions.

ACKNOWLEDGMENTS

We thank the editor and two anonymous reviewers for helping us explicit and clarify the contribution we hope to make with the present study. This work was supported by the Marcus and Amalia Wallenberg Foundation, under Grant [MAW 2015:0075].

REFERENCES

- Adam, Barbara. 1995. *Timewatch: The Social Analysis of Time*. Cambridge: Polity Press.
- Ayaß, Ruth and Cornelia Gerhardt, Cornelia. 2012. *The Appropriation of Media in Everyday Life*. Amsterdam: John Benjamins.
- Beach, Wayne. 1993. “Transitional Regularities for ‘Casual’ ‘Okay’ Usages.” *Journal of Pragmatics* 19:325–352.
- Becker, Howard. 1974. Art as Collective Action. *American Sociological Review*, 39(6):767–776.

- Bergson, Henri. [1889] 2001. *Time and Free Will: An Essay on the Immediate Data of Consciousness*. Mineola, NY: Dover Publications.
- Bhatia, Aditi. 2018. "Interdiscursive Performance in Digital Professions: The Case of YouTube Tutorials." *Journal of Pragmatics* 124:106–120.
- Brooker, Phil and Wes Sharrock. 2016. "Collaborative Music-Making with Digital Audio Workstations: The 'Nth Member' as a Heuristic Device for Understanding the Role of Technologies in Audio Composition." *Symbolic Interaction* 39(3):463–483.
- Broth, Mathias, Lorenza Mondada, and Eric Laurier, (eds). 2014. *Studies of Video Practices: Video at Work*. London: Routledge.
- Brown, Barry and Eric Laurier. 2005. "Maps and Journeys: An Ethnomethodological Investigation." *Cartographica* 40(3):17–33.
- Brown, Barry, Kenton O'Hara, Moira McGregor, and Donald McMillan. 2018. "Text in Talk: Lightweight Messages in co-Present Interaction." *ACM Transactions on Computer-Human Interaction* 24(6) Article 42: 1–25.
- Couper-Kuhlen, Elizabeth. 2014. "What Does Grammar Tell us about Action?" *Pragmatics* 24(3):623–647.
- Deppermann, Arnulf. 2019. "Intersubjectivity and Other Grounds for Action-Coordination in an Environment of Restricted Interaction: Coordinating with Oncoming Traffic when Passing an Obstacle." *Language & Communication* 65:22–40.
- Deppermann, Arnulf and Susanne Günthner. 2015. *Temporality in Interaction*. Amsterdam: John Benjamins.
- Deppermann, Arnulf and Jürgen Streeck. 2018. *Time in Embodied Interaction*. Amsterdam: John Benjamins.
- Dynel, Marta. 2014. "Participation Framework Underlying YouTube Interaction." *Journal of Pragmatics* 73:37–52.
- Engström, Arvid, Oskar Juhlin, and Mathias Broth. 2010. "Temporal Hybridity: Mixing Live Video Footage with Instant Replay in Real Time." CHI'10: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, April 10–15, Atlanta, Georgia, USA.
- Evans, Bryn and Oskar Lindwall. 2020. "Show Them or Involve Them? Two Organizations of Embodied Instruction." *Research on Language and Social Interaction* 53(2):223–246.
- Flaherty, Michael G. 2011. *The Textures of Time: Agency and Temporal Experience*. Philadelphia, PA: Temple University Press.
- Flaherty, Michael G. 2016. "Let's Study Time and Technology Rather than Editorialize." *Symbolic Interaction* 39(3):504–506.
- Frobenius, Maximiliane. 2011. "Beginning a Monologue: The Opening Sequence of Video Blogs." *Journal of Pragmatics* 43:814–827.
- Frobenius, Maximiliane. 2013. "Pointing Gestures in Video Blogs." *Text & Talk* 33(1):1–23.
- García-Rapp, Florencia. 2017. "'Come Join and Let's BOND': Authenticity and Legitimacy Building on YouTube's Beauty Community." *Journal of Media Practice* 18(2–3):120–137.
- Garfinkel, Harold. 2002. *Ethnomethodology's Program: Working out Durkheim's Aphorism*. Lanham, MD: Rowman & Littlefield.
- Goodwin, Charles. 1980. "Restarts, Pauses, and the Achievement of a State of Mutual Gaze at Turn-Beginning." *Sociological Inquiry* 50(3–4):272–302.
- Goodwin, Charles. 2002. "Time in Action." *Current Anthropology* 34(4):19–35.
- Goodwin, Charles. 2007. "Environmentally Coupled Gestures." Pp. 195–212 in *Gesture and the Dynamic Dimension of Language*, edited by, edited by S. Duncan, J. Cassell, and E. Levy. Amsterdam: John Benjamins.
- Gregg, Melissa. 2018. *Counterproductive: Time Management in the Knowledge Economy*. Durham, NC: Duke University Press.
- Haddington, Pentti, Tiina Keisanen, Lorenza Mondada, and Maurice Nevile. 2014a. *Multiactivity in Social Interaction: Beyond Multitasking*. Amsterdam: John Benjamins.

- Haddington, Pentti, Tiina Keisanen, Lorenza Mondada, and Maurice Nevile. 2014b. "Towards Multiactivity as a Social and Interactional Phenomenon." Pp. 3–32 in *Multiactivity in Social Interaction: Beyond Multitasking*, edited by, edited by P. Haddington, T. Keisanen, L. Mondada, and M. Nevile. John Benjamins: Amsterdam.
- Heath, Christian. 1986. *Body Movements and Speech in Interaction*. Cambridge: Cambridge University Press.
- Heath, Christian and Paul Luff. 2000. *Technology in Action*. Cambridge: Cambridge University Press.
- Heinemann, Trine and Ragna Lisa Möller. 2015. "The Virtual Accomplishment of Knitting: How Novice Knitters Follow Instructions when Using a Video Tutorial." *Learning, Culture and Social Interaction* 8:25–47.
- Jefferson, Gail. 2004. "Glossary of Transcript Symbols with an Introduction." Pp. 13–31 in *Conversation Analysis: Studies from the First Generation*, edited by, edited by G.H. Lerner. Philadelphia, PA: John Benjamins.
- Julien, Chris. 2019. "Couch Revisited: A Theoretical Treatment of the Information-Technological Media of Imgur, Reddit, and Twitter." *Symbolic Interaction* 42(1):46–69.
- Keevallik, Leelo. 2015. "Coordinating the Temporalities of Talk and Dance." Pp. 309–336 in *Temporality in Interaction*, edited by, edited by A. Deppermann and S. Gunthner. Amsterdam: John Benjamins.
- Kruse, Nathan B. and Kari K. Veblen. 2012. "Music Teaching and Learning Online: Considering YouTube Instructional Videos." *Journal of Music, Technology and Education* 5(1):77–87.
- Laurier, Eric and Barry Brown. 2014. "The Mediated Work of Imagination in Film Editing. Proposals, Suggestions, Reiterations, Directions, and Other Ways of Producing Possible Sequences." Pp. 237–261 in *Studies of Video Practices: Video at Work*, edited by, edited by M. Broth, L. Mondada, and E. Laurier. London: Routledge.
- Laurier, Eric, Barry Brown, and Moira McGregor. 2015. "Mediated Pedestrian Mobility: Walking and the Map App." *Mobilities* 11(1):117–134.
- Licoppe, Christian and Julien Morel. 2014. "Mundane Video Directors in Interaction. Showing One's Environment in Skype and Mobile Video Calls." Pp. 135–160 in *Studies of Video Practices: Video at Work*, edited by, edited by M. Broth, L. Mondada, and E. Laurier. London: Routledge.
- Licoppe, Christian and Sylvaine Tuncer. 2014. "Attending to a Summons and Putting Other Activities 'on Hold': Multiactivity as a Recognisable Interactional Accomplishment." Pp. 167–190 in *Multiactivity in Social Interaction: Beyond Multitasking*, edited by, edited by P. Haddington, T. Keisanen, L. Mondada, and M. Nevile. Amsterdam: John Benjamins.
- Lindwall, Oskar and Anna Ekström. 2012. "Instruction-in-Interaction: The Teaching and Learning of a Manual Skill." *Human Studies* 35:27–49.
- Lindwall, Oskar, Gustav Lymer, and Christian Greiffenhagen. 2015. "The Sequential Analysis of Instruction." Pp. 142–157 in *The Handbook of Classroom Discourse and Interaction*, edited by, edited by N. Markee. Malden, MA: John Wiley & Sons.
- Livingston, Eric. 2008. *Ethnographies of Reason*. Farnham, UK: Ashgate.
- Luff, Paul and Christian Heath. 2002. "Broadcast Talk: Technologically Mediated Action in a Complex Setting." *Research on Language and Social Interaction* 35(3):337–366.
- Luff, Paul, Jon Hindmarsh, and Christian Heath. 2000. *Workplace Studies. Recovering Work Practices and Informing System Design*. Cambridge: Cambridge University Press.
- McIlvenny, Paul. 2011. "Video Interventions in 'Everyday Life': Semiotic and Spatial Practices of Embedded Video as a Therapeutic Tool in Reality TV Parenting Programmes." *Social Semiotics* 21(2):259–288.
- Melander, Helen and Fritjof Sahlström. 2009. "Learning to Fly — The Progressive Development of Situation Awareness." *Scandinavian Journal of Educational Research* 53:151–166.

- Mondada, Lorenza. 2014. "Cooking Instructions and the Shaping of Things in the Kitchen." Pp. 199–226 in *Interacting with Objects: Language, Materiality, and Social Activity*, edited by, edited by M. Nevile, P. Haddington, T. Heinemann, and M. Rauniomaa. Amsterdam: John Benjamins.
- Mondada, Lorenza. 2018. "Multiple Temporalities of Language and Body in Interaction: Challenges for Transcribing Multimodality." *Research on Language and Social Interaction* 51(1):85–106.
- Nevile, Maurice. 2007. "Action in Time: Ensuring Timeliness for Collaborative Work in the Airline Cockpit." *Language in Society* 36:233–257.
- Nishizaka, Aug and Masafumi Sunaga. 2015. "Conversing while Massaging: Multidimensional Asymmetries of Multiple Activities in Interaction." *Research on Language and Social Interaction* 48(2):200–229.
- Perry, Marc, Mathias Broth, Arvid Engström, and Oskar Juhlin. 2019. "Visual Narrative and Temporal Relevance: Segueing Instant Replay into Live Broadcast TV." *Symbolic Interaction* 42(1):98–126.
- Råman, Joonas and Pentti Haddington. 2018. "Demonstrations in Sports Training: Communicating a Technique through Parsing and the Return-Practice in the Budo Class." *Multimodal Communication* 7(2):1–19. <https://doi.org/10.1515/mc-2018-0001>.
- Rauniomaa, Mirka, Pentti Haddington, Helen Melander, Anne-Danièle Gazin, Mathias Broth, Jakob Cromdal, Lena Levin, and Paul McIlvenny. 2018. "Parsing Tasks for the Mobile Novice in Real Time: Orientations to the Learner's Actions and to Spatial and Temporal Constraints in Instructions-on-the-Move." *Journal of Pragmatics* 128:30–52.
- Sacks, Harvey, Emanuel A. Schegloff, and Gail Jefferson. 1974. "A Simplest Systematics for the Organization of Turn-Taking for Conversation." *Language* 50(4):696–735.
- Schegloff, Emmanuel A. 1987. "Recycled Turn Beginnings: A Precise Mechanism in Conversation's Turn-Taking Organization." Pp. 70–85 in *Talk and Social Organization*, edited by, edited by G. Button and J.R.E. Lee. Clevedon, UK: Multilingual Matters, Ltd.
- de Stefani, Elwys and Anne-Danièle Gazin. 2014. "Instructional Sequences in Driving Lessons: Mobile Participants and the Temporal and Sequential Organization of Actions." *Journal of Pragmatics* 65:63–79.
- de Stefani, Elwys and Anne-Sylvie Horlacher. 2018. "Mundane Talk at Work: Multiactivity in Interactions between Professionals and their Clientele." *Discourse Studies* 20(2):221–245.
- Strangelove, Michael. 2010. *Watching YouTube: Extraordinary Videos by Ordinary People*. Toronto: University of Toronto Press.
- Streeck, Jürgen, Charles Goodwin, and Curtis LeBaron. 2011. *Embodied Interaction: Language and Body in the Material World*. Cambridge: Cambridge University Press.
- Suchman, Lucy. 2007. *Human-Machine Reconfigurations. Plans and Situated Actions*. 2nd ed. Cambridge: Cambridge University Press.
- Szymanski, Margaret. 1999. "Re-Engaging and Dis-Engaging Talk in Activity." *Language in Society* 28(1):1–23.
- Tuncer, Sylvaine, Barry Brown and Oskar Lindwall. 2020. "On Pause: How Online Instructional Videos are Used to Achieve Practical Tasks." *CHI'20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, April 2020, pp. 1–12.

ABOUT THE CONTRIBUTOR(S)

Sylvaine Tuncer is a sociologist, researcher at the Royal Institute of Technology (KTH, Stockholm, Sweden) in the Robotics, Perception and Learning division. Her approach is based in conversation analysis, ethnomethodology, and ethnography. She studies the emergence of social organization through interactions, technologies, and artifacts. Her empirical research spans from workplace

studies to mobilities in the public space, through human-robot interactions and video-mediated communication. She has published in various specialized venues and formats, from linguistics journals to human–computer interaction conference proceedings, through methodological contributions.

Oskar Lindwall is an Associate Professor at the Department of Applied Information Technology, University of Gothenburg, Sweden. His research interests include the use of technology in higher education, video research in the learning sciences, ethnomethodology, and conversation analysis. He has conducted research on demonstrations in dental education, critique session in architectural education, practices of instruction in sports coaching, the teaching and learning of crafts, and simulations in healthcare and maritime education. He is currently the principal investigator in the project “The sensitive nature of feedback as an interactional phenomenon” funded by Riksbankens Jubileumsfond.

Barry Brown is a research professor at the University of Stockholm, where he helps to runs the STIR research group. He previously (2011–2017) worked as the research director of the Mobile Life research center and as an associate professor in the Department of Communication at UCSD (2007–2011). His two most recent books, focusing on how to research the use of digital technology, and the study and design of leisure technologies, have been published by Sage and MIT Press.