

Title: Doing Inspecting in Interaction: Seeing the Physiognomy of an Object

Author: Aug Nishizaka

Affiliation: Chiba University, Japan

E-mail address: augnish@chiba-u.jp

Funding:

This work was supported by the Japan Society for the Promotion of Science under the Grant-in-Aid for Scientific Research (C) (20K02131).

Declaration of competing interest:

The author declares that there is not conflict of interest.

Biographical note:

Aug Nishizaka is a Professor of Sociology at Chiba University. His research has been concerned with the interactional organization of visual and tactile perception. His publications include: 'Seeing what one sees: Perception, emotion and activity,' *Mind, Culture, and Activity* (2000), 'The perceived body and embodied vision in interaction,' *Mind, Culture, and Activity* (2017), and 'Appearance and action: Sequential organization of instructions in Japanese calligraphy lessons,' *Research on Language and Social Interaction* (2020).

Doing Inspecting in Interaction: Seeing the Physiognomy of an Object

Abstract

This study explores the practice of doing inspecting an object, more specifically, the practice of leaning over an object that has been seen in a certain way. It offers a single-case analysis of two segments in which doing inspecting is done in an enhanced way, accompanied by touching the object. It argues that seeing the details of an object is done not necessarily to collect detailed information about the object but as a constitutive part of the ascribability of a specific action to the whole of the viewer's concurrent verbal and other behavior. Seeing the details of an object is seeing the object in an entirely new fashion. Following the empirical analysis, its implications for some aspects of perception (multimodality of perception, perspectives, and the unity of the body) will be discussed. Data are in Japanese with English translation.

Keywords: doing inspecting; seeing details; conversation analysis; multimodal perception; action

Introduction

The purpose of this study is to demonstrate that *seeing the details* of an object is a constitutive part of the ascribability of a specific action to the whole of the viewer's concurrent verbal and other behavior¹; it is not done specifically to collect more information about the object. In other words, seeing the details is not necessarily a condition for an action's successful performance but is constitutive of what type of action (e.g., a request or question) is ascribable to the whole of the viewer's concurrent behavior. To demonstrate this, we examine cases in which a participant specifically inspects an object while touching it. Visual perception has been one of the focuses of social interaction studies. In particular, visual orientation (i.e., where the participants look) has been one of the most developed areas since Kendon's (1967) seminal work on eye gazing. Since Kendon's work, many scholars have investigated how gazing functions in turn allocation (see Auer, 2018, 2021; Stivers & Rossano, 2010) and sequence organization (Rossano, 2012). Many studies have also shown that gazing contributes to the organization of "participation frameworks" (Goffman, 1979). In particular, a series of C. Goodwin's and M. H. Goodwin's studies has explored various manners in which participants' orientations to one another and objects in the environment—orientations differently exhibited in different body parts, postures, and talk—are coordinated to organize a specific action (C. Goodwin, 1981, 1984, 2007; M. H. Goodwin, 1997; Goodwin & Goodwin, 2004; see also Nishizaka, 2013, 2014b). All these studies have indicated that what each participant sees (i.e., visually perceives) in interactions is available to other participants and serves as a crucial resource for the organization of a joint activity.

Some studies have also addressed the organization of how participants see particular objects (which may be referred to as a “visual experience”) since a series of Goodwin’s (1994, 1996) seminal work. Goodwin (1994) explored the organization of what he called “professional vision”; using pointing gestures combined with classificatory words within a specific temporal and spatial arrangement of multiple bodies perceptually restructures how an object is to be seen (see also Heath & von Lehn, 2004; Nishizaka, 2011, 2014a). This “how” of seeing is necessary to perform an action appropriately for the practical purposes of the participants’ profession. Goodwin (1996) showed that a designedly unspecified referential term (such as “a problem”) may perceptually restructure the current visual scene as specifying the term (such as the scene becoming visible as “problematic”) (see also Nishizaka, 2000). Goodwin and Goodwin (1996) demonstrated that how participants see particular objects depends on the activity they are engaging in. For example, workers at an airport see not only the difference between airplanes with aircraft-identification numbers on their bodies but also the difference between flights with specific destinations. This specific seeing is specifically relevant to their activities (such as loading luggage into airplanes).

Nishizaka’s (2020a) study of interactions between a Japanese calligraphy master and his students focused on the interactional organization of the particulars of visual experience. He showed that the master, using various highlighting practices, organizes seeing the drawing action of stroking or sweeping as a constituent part of the appearance of a drawn character and that this specific seeing is consequential to the sequential organization of the current instructing activity. This seeing reminds us of what some philosophers have observed under the rubric of “synesthesia.” The drawing action is not physically present before the participants’ eyes, but they see it in a way

similar to seeing the rigidity of glass without touching it (Merleau-Ponty, 2012) or perceiving the “atmosphere” of a scene at the moment that the scene opens itself before the viewer who moves into it (Böhme, 2013, p. 95).

Touching another participant’s body is one of the recently widely studied areas in interaction studies (see Cekaite & Mondada, 2020; Burdelski et al., 2020, for overviews). While the previous studies have well documented the various interactional functions of various types of human-to-human touches, Nishizaka (2020b) focused on the interactional organization of the details of tactile experience. In his analysis of “guided touch” in midwifery settings (i.e., a midwife guiding a pregnant woman’s hand to tactilely feel the fetus), he demonstrated that publicly witnessable subtle differences in how the participants feel the fetus (and they feel each other feel it) are consequential to the sequential development of the interaction.

In real life, we encounter the world multimodally. As the aforementioned philosophers emphasized, “synesthesia” is not a special experience (much less pathological one) but rather the most fundamental experience (see also Waldenfels, 2000). In addition, when we see or touch something, the seeing or touching cannot be independent of our proprioception (e.g., sensing where our own limbs are positioned or moving and sensing how we are moving our body parts or entire body). Furthermore, we often see an object while touching it. How integrated multimodal or multisensory perception is organized in interaction has recently been one of the central issues in interaction studies. Nishizaka (2007, 2010, 2020b) addressed this issue by examining interactions between a pregnant woman and a medical professional who examines an “object” inside the other’s body. More recently, Mondada (2021) demonstrated the centrality of such multisensory (or multisensorial, to use her term) perception of an

object, focusing on tasting. In tasting, not only are gustatory, olfactory, and tactile perceptions indiscriminably fused (see Gibson, 1966); tasting, looking at, and smelling cheese were organized in their mutual dependence and in their conjuncture with various embodied behaviors. She showed that tasting is interactionally organized as a constituent part of the witnessable order of distinct activities.

In psychology, multimodal perception has been intensively investigated (see Soto-Faraco et al., 2014, for an overview of studies of multimodal perception in the real world). The following two studies, for example, are relevant to this study. Higashiyama and Adachi (2006) showed that proprioception—the sense of one’s own postures, in particular—influences the perception of the size and distance of objects. Shutz and Lipscomb (2006) found that how a marimba player strokes their mallets influences how long a viewer of the performance hears the sound persist; the result is contradictory to the previous studies, which had found that audition influences vision rather than the other way around. These studies suggest the possibility that perception is deeply embedded in the perceiver’s bodily configuration and—more relevantly to this study—in the activity in which the perceiver is engaging.

This study explores the relationship between perception and action, by examining naturally occurring interactions, focusing on doing inspecting an object in interaction. Specifically, it addresses the following questions: What is accomplished by doing inspecting an object? How are the phenomenological particulars of how one perceives the object during inspection consequential to the sequential development of the ongoing interaction? However, the main purpose of this study is not to explicate the generic properties of the focal practice—doing inspecting; rather, it uses the practice as a lens through which several aspects of seeing an object in interaction are explored by

addressing these questions. We demonstrate that seeing the details of an object is a constitutive part of the ascribability of a specific action to the whole of the viewer's concurrent behavior and is not necessarily done to collect detailed information about the object in order to perform the action better. What action type is ascribable to the whole configuration of the viewer's concurrent behavior may vary according to whether doing inspecting is part of it. In what follows, we first specify the phenomenon on which this study focuses by situating it in the literature on inspecting and by presenting a relatively simple case of doing visually inspecting to provide a first sense of the phenomenon. In the main body of this study, we analyze the most illuminating example and then analyze a supplemental example. In these examples, doing inspecting is done in an enhanced manner that involves touching the object in certain ways; they provide perspicuous settings in which seeing is shown to be an essentially "intercorporeal" (Merleau-Ponty, 1960) phenomenon. We then discuss the implications of the analyses to explore several aspects of perception. We argue that perception is essentially multimodal, that the perceiver's perspective essentially involves others' perspectives, and that the unity of the body is organized and maintained in performing actions.

Doing inspecting in interaction

Mondada (2018) explored the interactional organization of tasting activity—that is, customers' activity of inspecting the taste of cheese in cheese shops—and found a generic organization of such inspection. Partially drawing on Mondada's study, Mortensen and Wagner (2019) explicated the structure of what they called "inspection sequences"; they observed that how these "sequences" are initiated and completed has a

generic pattern and that this pattern can be adapted to local contingencies, including the difference in the materiality of the objects being inspected. In these studies, inspecting is a focal activity for the inspector, who is doing concentrating on inspecting by doing it silently, while the inspecting is embedded in and tailored to its sequential and praxiological environment. Mortensen and Wagner also documented cases in which inspecting is interactionally organized as an individual activity. In contrast, the inspecting that this study addresses is incorporated into the inspector's ongoing action accompanied by talk and vocal behavior.

Streeck's (1996) study of inspecting a cookie, a product of a company, argued that the inspection transforms the symbolic nature of the cookie under inspection; for example, inspecting it (by looking at, touching, and biting it), carefully orchestrated with the running report of it, reconstructs the cookie as a *sample* of the product. Streeck's study is relevant to this study in that it addresses not only the structure of the inspecting activity but also how the multimodal experience of the inspected object is organized in the very practice of inspecting (see also Mondada, 2019, for a similar transformation of the nature of an object).

The simplest case of visual inspection (without touch) of the type that this study addresses is found in Nishizaka (2000, 2018). In Excerpt 1, taken from a lesson on using a word processor on a DOS/V machine, the teacher instructs the student to input a "half-sized" (normal-sized) "IBM." Using a Japanese word processor at that time (in the 1990s), to obtain a half-sized "IBM," one had to first input three full-sized characters, then press function key 9 (F9) to obtain three full-sized alphabets, and finally press F8; the order was crucial for the completion of the task. In following the teacher's instruction, the student looks down at the keyboard ("K") and then follows the

instruction, looking at the screen (“M”) and the keyboard alternately (data not shown). After the student presses a function key (line 01), some trouble surfaces (lines 02–03). Then, however, the student makes a second attempt (line04) and finally marks the completion of the task by raising her body (line 05). In response, the teacher produces admiration and evaluates the student’s performance while raising his body (line 06). The focus of this study is on the teacher’s conduct in line 02. See Appendix for symbols employed in the excerpts.

(1) [Nishizaka, 2000, 2018]

01 / (6 . 4) / (0 . 8)
 stu.g: KKKKKKKKKKKkmMMMmkKK
 stu: |presses |moves r.h. onto
 3 keys-->|a function key

02 TEA:→ nn (.) | (.) | °are°?
 Yeah | | °O?°
 stu.g: kmMMMMMMMMMMMMMm
 stu: |presses a function key
 tea: |thrusts his upper body
 toward the screen

03 STU: a r e ?
 O?
 stu.g: kKKkM

Fig. 2
 ↓

04 (0 . 8) / (0 . 4) / (.)
 stu.g: MMMMMMMMMMMMMMMMMM->>
 stu: |presses |presses
 a key a key

05 STU: | (° [| ne °) |
 06 TEA: | [a / aa aa : : ↓ : . (0 . 2) soo des' ne
 Oh:: | **That's the way.**
 stu: |raises upper |
 body and |
 nods twice ->|
 tea: |raises upper body

07 TEA: >dakara< (0 . 6) ikinari (0 . 8) hachi ban o
 so first eight number P
 08 oshi tara ...
 press if
 So, if you press number 8 first, ...



Figure 1
TEA points at the screen when providing the instruction (*before Excerpt 1*).



Figure 2
TEA is doing inspecting the details of what is visible on the screen.

In line 08, the teacher acknowledges the student's pressing of the first function key (although at this moment he most probably sees the student has pressed the wrong key), but immediately after the student presses the second function key, he is doing noticing something wrong on the screen by producing a token indicating that something unexpected has appeared (*are? "o?"*) and thrusting his upper body toward the screen (compare the positions of his upper body in Figures 1 and 2). In this fashion, the teacher is here doing inspecting what happens on the screen. This is the practice on which the present study focuses. In Excerpt 1, the teacher's doing inspecting the details on the screen accomplishes at least two things: First, it provides the groundedness of the explanation of the incorrect operation in what the teacher has seen; because of the doing inspecting, the explanation that the teacher offers (lines 13–14) is hearable as being based on the details that he saw when the student's attempt was unsuccessful. Second, the teacher encourages the student to fix the situation herself via his doing displaying that the solution is not immediately available to him. Thus, inspecting the details appears to be a constitutive part of the ascribability of a specific action, namely, encouraging continuation, not only (or rather than) collecting some relevant information. We will further develop this point by examining more complex examples involving touch as well as vision.

Data

In what follows, we focus on the specific conduct of doing inspecting: bringing one's face (or eyes) close to a specific (or specifiable) object by leaning toward the object. We went through the Corpus of Everyday Japanese Conversation (Koiso et al. 2016) to search for such conduct. We began by mechanically collecting all the examples of this conduct, but we excluded cases in which the object being inspected was not (or may not have been) adequately visible to the inspector, such as when one brings one's face closer to a smartphone to read small letters or closer to the floor to search for something that one has dropped. In the end, we had ten cases in which someone was doing seeing the *details* of an object that had been visible to them in certain ways, as the teacher is doing in line 08 of Excerpt 1. In this study, we examine two of these cases, in which the participants are doing inspecting in an enhanced way—that is, inspecting while touching the object or while holding the object stationary with a hand—and what the inspecting accomplishes is transparent to a certain degree. We transcribed all details of the interactions in these cases, using the transcription system developed by Jefferson (2004). We use conversation analysis (Sacks, 1992; Schegloff, 2007) to analyze them. Conversation analysis transcribes and examines the details of actual interactions and documents the participants' orientations (such as what type of action *they* ascribe to their utterances) exhibited in the details of the interactions. We ground our analytic claims in these orientations.

Action accomplished by doing inspecting

The next example (Excerpt 2) is taken from an interaction between the owner (OW) of a consignment shop and an accessory designer (DS) who has brought her products into the shop. Before the excerpt, the owner has told the designer that to use the shop to sell her products, two options are available for how to pay the consignment fee: (1) the designer could pay a determined amount as a fee to the shop or (2) she could let the shop take a certain percentage of the prices of the sold products. In lines 01 and 04, the designer, indicating that she has currently set the prices of her products low, suggests that she is ready to add some amounts to the prices, which the owner could take as fees. In response, the owner first states that she can see, from the appearance of the products, what the designer has indicated (line 07) and then, after proceeding to but self-interrupting the positive evaluation of the products (with the self-interrupted em-phasizers *zu-* [*zuibun* “fairly”] and *sugoi-* [*sugoi* “very”]), which are hearable as the incipient evaluation [see Goodwin & Goodwin, 1987]), inquires whether the designer can rewrite the prices (lines 07–09); thereby, the owner exhibits her understanding that the designer has suggested the changing of the prices. The focus here is on the action implemented by this question.

(2) [CEJC K002_004]
01 DS: *etto::* *ima kono tsuke te'ru nedan teyuuno wa:* *Γ:*
well | now this write have price about P |
Well, regarding these current prices that I have written
down, | *nn* | *nn*
02 OW: | *nn* | *nn*
Yeah | **Yeah**
03 DS: *.hhhh/(0.8)*
04 DS: *ee::to (.) chotto yasumeno settee::::* *Γ:::: ni shi₇te'run'*
well kind.of low set | P have.done
Well, (.) ((I)) have set ((them)) kind of low, |
05 OW: | *nn nn nn nn*
Yeah, yeah, yeah, yeah
06 DS: *=↓desu ↑ne*
POL P
you know.

07 OW:→ *soo deshoo ne* *zu- sugoi- .HH* *tatoeba:* *kore zenbu jaa=*
so guess P | for.example this |all then
I can imagine. **Fair- Ver- .HH** **For instance, then, all these,**

08 DS: *~n:n* *~n*
Mmhmm **Mm**

09 OW:→ *=kakikaete morau koto toka deki mas'?*
rewrite take thing like can POL
can ((you)) rewrite them or something?

10 *<ima .hh ano:: kokoni >tatoeba< ni sen roppyaku=*
| now | uh here for.example 2000 600
|<now .hh |Uhm here, for instance, changing to 2,600

11 DS: *~a zenzen sore wa:: (.) deki masu:* *~*
oh absolutely that P can AUX.POL
Oh, if that is it, ((I)) can absolutely do it.

12 OW: *=en ni suru to ka:?*
yen P make like.that
yen, or like this.

13 DS: *hai:*
Yes.

14 DS: *ha::i*
Yes.

15 OW: *<sonna katachi: de::>*
In this way

16 DS: *sore wa deki masu :*
that P can AUX.POL
If that is it, ((I)) could do it.

17 OW: *yari mashoo ka*
do AUX.POL IT
shall ((we)) do ((it))?

In her inquiry, the owner uses the “can you...?” (*deki mas'?*) format (line 9); this format may be used to make a request or a proposal. However, just after the first possible completion of the owner’s current turn (i.e., just after the utterance of *deki mas'?*), the designer responds with the repetition of *deki masu*, preceded by a grammatical object (*sore* “it”) with the contrastive marker *wa* and emphasized by the modifier *zenzen* (“absolutely”). First, the repetition of *deki masu* appears to be descriptive of the possibility or ability rather than accepting of a request or proposal (she could have said *ii desu yo* or *wakari mashita*, “all right” or “I will,” as the acceptance of a request or proposal).² Second, the contrastive marker *wa* (of *sore wa*, translated as “if that is it”) indicates that something more than what the designer currently states will ensue. Together, the designer appears to claim that it is (absolutely) possible for her to rewrite (if this is all that is inquired about). In fact, after the exchange

in the excerpt, the designer reveals that she currently has extra tags with her, thereby indicating that it is indeed possible for her to rewrite right now. In other words, this response of the designer is hearable as exhibiting her understanding that the owner’s “can-you” question was inquiring about the possibility of the designer rewriting the prices rather than requesting that she do so and was a question *preliminary* to the arrangement of commission fees to come next. In fact, in lines 15 and 17 (“In this way, shall ((we)) do ((it))?”), based on the designer’s assertion that she can rewrite the prices, the owner appears to proceed to suggest explicitly what was implied by the designer’s first statement (“I have set the prices low”; lines 01 and 04).

Now, if this hearing is correct, the following issue should be addressed: What is the mechanism by which the owner’s “can-you” question can be an inquiry about the possibility rather than a vehicle for a request or proposal? We will now address this question by examining embodied behavior. The following excerpt (2a) is a detailed transcript of lines 01–02 and 07–11 of Excerpt 2.

(2a)

		fig. 3		
		↓		
01 DS:	<i>eetto::</i> <u>ima</u> <i>kono tsuke te'ru nedan te/yuuno wa:</i>			
	well now this write have price like P			
	Well, regarding these current prices that I have written			
	down,			
02 OW:	<u>nn</u>			<u>nn</u>
	Yeah			Yeah
ds:				touches a tag

((4 lines omitted))

07 OW: *soo deshoo ne* [zu- sugoi- / .HH |↑tatoe|ba: kore [zenbu jaa=
 so guess P | | |for.example these|all then
I can imagine. Fair- Ver- .HH For instance, then, all these,
 08 DS: ^{L_n:n} Mmhmm | | | ^{L_n} Mm
 ow: |abruptly leans toward the table
 while extending l.h. and touches
 a tag with fingertips
 ds: |withdraws r.h.
 ds: |steps backward while
 looking at what ow
 touches

fig. 4

↓
 09 OW: =ka/kikaete morau koto toka deki mas'?
can ((you)) rewrite them or something?
 ow: |takes another tag
 with fingertips

fig. 5

↓
 10 [ima .hh [ano:: kokoni >tatoeba< ni↑sen roppyaku en ni=
 [now .hh [Uhm here, for instance, changing to 2600 yen,
 11 DS: [a zenzen [sore wa:: (.) deki ma↓su:↓
Oh, if that is it, ((I)) can absolutely do it.



Figure 3
 DS touches a tag with her upper body
 torqued toward OW.



Figure 4
 OW leans toward a tag on the table
 and takes it with her fingertips. DS
 has slightly stepped back and watches
 OW's inspection.

When taking a sharp in-breath in line 07 just before the production of the question, the owner abruptly leans over the table and extends her left hand, the hand closer to the designer, toward the tag that the designer touches (compare the owner's postures in Figures 3 and 4).³ When uttering a deictic term *kore* ("these"; line 07), the owner touches a tag and then proceeds to touch another tag with two fingertips. With

this exhibited intent orientation to the tags, particularly with her face brought closer to the tag (Figure 4), the owner appears to be doing inspecting the details of the price tags. Furthermore, in conjunction with her utterance of the indexical expression *kokoni* (“here”; line 10), while maintaining her inspecting posture, the owner touches another tag with the tip of her index finger with the palm oriented downward (Figure 5); she appears to be doing pointing at a specific part of the object, thereby doing inspecting the conditions of the tiny spaces for the prospective action to be done—that is, doing inspecting the tags to see whether they will allow for rewriting the prices on the tags. Thus, the “can-you” question becomes hearable as concerning the possibility of rewriting prices in these spaces.⁴

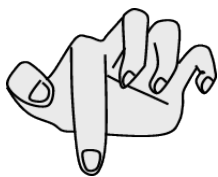


Figure 5
Pointing with the palm oriented downward

In sum, the owner’s practice of doing inspecting the details of the objects makes *the seeing of the conditions of the tiny spaces for the prices to be rewritten* ascribable to the owner. This ascribable seeing makes the owner’s “can-you” question hearable as an inquiry about the possibility of rewriting the prices in these tiny spaces.⁵

Doing inspecting within a bodily configuration involving multiple bodies

Doing inspecting the details of the price tags is only possible within the bodily configuration in which the owner’s talk and embodied orientations (including gaze

direction, movement of arms and upper body, body orientation, etc.) are juxtaposed with each other. This bodily configuration is also placed within a temporal and spatial arrangement of the two (the owner's and designer's) bodies in which the designer is doing seeing the owner's practice of doing inspecting (see Goodwin, 2017, chapter 21). Specifically, when the owner leans over the table (line 07), the designer withdraws her right hand and steps backward while maintaining her gaze toward the table (Figure 4). In this fashion, the designer appears to secure adequate space for the owner's doing inspecting and to position herself as one who, commanding an overall view of the owner's inspecting, observes it without manipulating things on the table herself. Moreover, the owner, who is right-handed, uses her left hand (closer to the designer) to touch the tags. This accomplishes at least two things: First, the touching is thus done at the center of their common focal field; second, the inspecting is thus done as more separated from the interaction with the co-participant, thereby doing concentrating on inspecting (Mondada, 2018; Mortensen & Wagner, 2019).

Thus, the owner's doing inspecting is interactionally accomplished in the spatial and temporal configuration of the participants' embodied orientations—the configuration that is composed of the system of the participants' orientations embodied by multiple bodies' postures and movements (e.g., orientation to the tags on the table embodied by the owner's leaning, extending a hand, gazing, and touching; orientations to the owner's orientation to the tags embodied by the designer's distancing herself from the table, posture, and gazing, etc.; Figure 4). In this bodily configuration, seeing the detailed conditions of the tiny spaces of the price tags becomes ascribable to the owner, and, thereby, the owner's action of inquiring about the possibility of rewriting is accomplished.

Doing inspecting as a disambiguating device

In both Excerpts 1 and 2, the utterances (“O?” and “Can you rewrite them?”) by the teacher and the owner, respectively, could form certain actions without doing inspecting. However, the utterance “O?” (Excerpt 1), marking the emergence of something unexpected, might imply a negative evaluation or even a criticism of the student’s maloperation. Doing inspecting might appear to disambiguate the utterance by doing searching for something wrong that may not be easily detectable and mitigating the student’s responsibility for the maloperation. If the “can-you” question by the owner of the shop (Excerpt 2) formed a request to rewrite the prices that the designer decided on, this might violate the designer’s right to finally decide the prices (Stevanovic & Peräkylä, 2012). Doing inspecting might appear to disambiguate the question by doing examining the detailed conditions of the tags. However, such disambiguation is not clearly oriented to by the participants in both examples; rather, to them (the participants), the utterances and embodied conduct are not separable units but form a single configuration (Goodwin, 2017). However, in this section, we suggest that doing inspecting *can be used* by participants as a disambiguating device (although it is not always so). We show that when a systematically ambiguous action contingently emerges, the removal of its ambiguity may be in order.

Excerpt 3 is taken from an interaction between five friends; they have gathered to celebrate the birthday of one of them (Natsu). The excerpt begins as Natsu (NAT) is opening her birthday present (two plates, which are part of a series of three bear-shaped plates). Figure 6 represents the arrangement of their seats. The sound *biri* in line 01 is a

conventional mimetic expression for the sound made when paper is torn. The focus here is on Mika's doing inspecting in line 07.

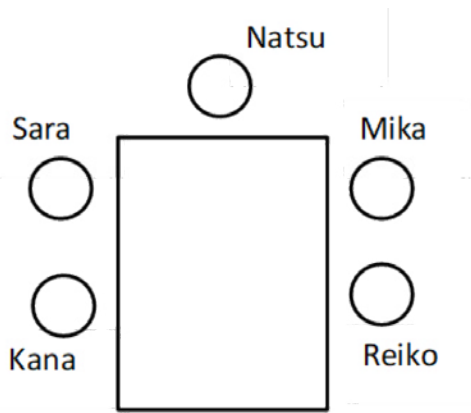


Figure 6
The arrangement of the participants' seats

```

(3) [CEJC C001_001]
01 NAT: biri h: ((mimetic expression))

02      (2.4)

03 ? : |.hhh
    rei: |leans forward

04 REI: |aa /kawa/i_i r::r: : /:
    Oh, |cute.
05 SAR: |aa /kawa-ii /: : /:
    Oh, |cute.
06 KAN: |>kawaii.<
        |Cute.

07 MIK:→ |heh hehh /.hh /hehh /.hh
    rei: ---->|holds the posture ----->>
    kan: |leans forward----->|raises her body
    mik: |leans |raises her body ->|extends r.h.
        forward->|
    nat: |turns the
        |plate to mik
    mik: |touches
        |the plate
    nat.g: mMMMMMMMMMM

08 NAT: rkore no ne: (chitchai ko:) o: motterun: desu yo:.
    {this P P smaller one P have POL P
    I have a smaller version of this.
09 MIK: L$kawa(ii$)
        $Cute.$
    nat.g: MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMmm
  
```

Fig. 7



Figure 7

Mika touches the plate that Natsu has turned to Mika. (The contours are distorted due to the use of a fisheye lens.)

After Natsu has opened the present, first Rei (line 03) and then Kana (line 04) lean forward toward the plate, thereby doing inspecting the present—a bear-shaped plate. They offer positive evaluations of it (“cute”; lines 04 and 06), for which their doing inspecting intelligibly provides perceptual grounds. Mika first leans toward the plate (first inspection) and, while raising her body back, begins to laugh (line 07), and then, while extending her right hand toward the plate, she brings her upper body closer to the plate again (second inspection, which is our focus here). Mika’s second inspection is enhanced, for Mika not only leans forward but also holds the plate stationary in front of her face (Figure 7). However, the enhanced inspecting only leads up to the same simple one-word evaluation that others have offered (“Cute.”; line 09); this suggests that the second inspection is not done to collect more information for a more detailed evaluation of the object (the plate). What does this enhanced doing inspecting accomplish, then?

Her second inspection is done in the context in which her first inspection leads up to the first evaluative behavior: laughing loudly (line 07). Note that the one-word evaluation (“cute”) offered by Rei, Sara, and Kana (lines 04–06) is a reactive expression, which is sequentially (proximally) positioned in relation to the object it is

reactive to. Mika's laughter is produced as an alternative reaction to the same object, and she may have reasons why she reacts to it in a different way than the others: Their reactions are sequentially organized not only relative to what they are reacting to but also in relation to each other. In other words, the reactions do not occur simultaneously but one after another (Excerpt 3a).

```
(3a) [lines 04–07]
04 REI: |aa [kawai_i r:::/r: : : <-- first
      |Oh, [cute
05 SAR: |aa kawa-ii //: : : <-- second
      |Oh, cute.
06 KAN: |>kawaii.< <-- third
      |Cute.
07 MIK: |heh hehh .hh hehh .hh <-- third
      |( 0 . 9 )
```

In this fashion, the fact that participants have the same evaluation is not merely coincident; the subsequent speakers are also agreeing with the prior speakers in the evaluation. This point is clearer in lines 11 and 12 of Excerpt 1 (Excerpt 1a).

```
(1a) [lines 05–06]
      Fig. 8   Fig. 9
      ↓       ↓
05 STU: | (° [ | ne°)
06 TEA: | [a /aa aa::↓:: (0.2) soo des' ne
      |Oh:: That's the way.
      stu: |raises upper body <-- first
      tea: |raises upper body <-- second
```



Figure 8
STU raises her upper body to mark the completion of the task.



Figure 9
TEA also raises his upper body.

As Nishizaka (2000) observed, while the teacher's raising his body sharply is

observably reactive to what he sees, by being produced after the student raises her body, his reaction also forms agreement with the student's seeing the correct result (Figures 8 and 9). However, Mika's third-positioned reaction may be already too distant from the appearance of the object to which it is reactive, and if she, as the third reactor, produced the same reaction at this sequential position, her reaction might have been taken as only superficially going along with the others without independently reacting to the object. Incidentally, Kana is also the third starter, and her quick articulation of the same evaluative word *kawaii* ("cute") without the noticing token *aa* ("oh")—thereby reducing the responsive nature of her evaluation—may be sensitive to this later start. Furthermore, the first two participants' one-word evaluations have their final sounds substantially extended so that they overlap each other. In this position, a lexical reaction may be difficult to hear adequately. Therefore, laughter, a non-lexical reaction that is more permeable to others' ongoing talk, may be most suitable here.

However, laughter is ambiguous as a reaction in that it implies either a positive (interestingness, enjoyability, etc.) or negative (stupidity, awkwardness, etc.) evaluation of the object it reacts to. The enhanced doing inspecting at this sequential position is understandable as being motivated by the aspects of the object the laughter is reactive to—namely, the aspects of the plate that deserve *further* inspecting for the unpacking of the meaning of the laughter. Doing inspecting in an enhanced manner in this context is also intelligible as doing enjoying looking at it, or at least it (re)constructs the object as worthy of inspecting or carefully examining. Thus, the enhanced doing inspecting disambiguates the laughter as an affective attitude toward a specifically interesting and enjoyable thing and incorporates the laughter into a rather strong positive stance toward it, although it ends up with the simplest positive evaluation ("cute").

Natsu's behavior is well geared toward Mika's conduct in some respects. When Mika begins to extend her hand toward the plate, Natsu gazes at her ("M") while turning the front of the plate toward her. Then, Natsu offers the reason for her request for this plate (line 08) simultaneously with Mika's offering a lexical evaluation (line 09). Offering the reason why she wanted it, which also conveys additional (potentially interesting) information about the plate (i.e., that it is part of a series), as well as revealing how much she likes (or is fascinated by and enjoys) the one she has, appears responsive to Mika's special interest displayed by her enhanced doing inspecting.

In this fashion, Mika's enhanced doing inspecting appears to be an operation on her ongoing laughter, the result of her first doing inspecting, rather than collecting more information about it to provide a more detailed evaluation. Thus, the enhanced doing inspecting may be used by Mika to disambiguate the meaning of the laughter.

In this section, we have indicated the possibility that doing inspecting can be used as a disambiguating device. We have grounded the possibility in a detailed analysis of an excerpt, by showing that a systematically ambiguous action (i.e., laughter with possible negative implications) is occasioned contingently and independently (of doing inspecting), and doing inspecting in the example appears to address such ambiguity. This possibility once again suggests that seeing is a constitutive part of the ascribability of a specific action to the whole of the viewer's concurrent behavior .

Seeing details in action

Doing inspecting is one of the witnessable practices for accomplishing the ascribability of "seeing details" to the actor/speaker. We have shown that seeing details does not

necessarily aim *specifically* at collecting detailed information. It is important to remember that saying something does not necessarily aim specifically at conveying information (Austin, 1962; Searle, 1969); saying something, or uttering words, is doing an action (greeting, promising, inviting, requesting, etc.). Similarly, seeing cannot be entirely equated with collecting information. Rather, seeing is a constitutive part of the ascribability of a specific action. If one saw things only to collect information, seeing details, which would collect more information, would always be favored. However, people do not always see details in order to collect more precise information (although, of course, seeing is always collecting some information, as speaking is always conveying some information). If seeing is a constitutive part of the ascribability of an action, the relevance of the details to be seen (or how much detail one should see here and now or “the relevant precision” [Drew, 2003] of seeing) depends on what type of action is accomplished with the seeing. For example, Wittgenstein (1953) asked the following question:

Am I inexact when I do not give our distance from the sun to the nearest foot, or tell a joiner the width of a table to the nearest thousandth of an inch? (para. 88)

Note that detailedness, calibration, and granularity, on the one hand, and exactness and precision, on the other, belong to different language games. Telling a joiner the width of a table to a tenth of an inch would be “inexact,” while giving our distance from the sun to a foot would be too exact (and would not make sense), although a tenth of an inch is a more detailed measurement than a foot (see Nishizaka, 2022). “‘Inexact’ is really a reproach, and ‘exact’ is praise” (Wittgenstein, 1953, para. 88; see also Coulter, 1991). The judgment of exactness and precision is a constituent part of the ascribability of an action (complaint, praise, or the like).

Seeing an object in its details may be seeing a physiognomy or “aspect” (Wittgenstein, 1953, 1980; see also Nishizaka, 2018, 2020a) of the object rather than obtaining more information about it. Seeing the details of an object may be seeing it totally differently, in the same way that seeing the well-known duck-rabbit figure as a rabbit is seeing the figure under a totally different physiognomy than seeing it as a duck. In other words, the details of an object reveal themselves as part of the organization of an action. This is what we demonstrated in the previous sections.

In the remainder of this section, we explore some implications of the demonstration: implications for (1) the multimodality of perception, (2) the perceiver’s perspective, and (3) the unity of the body.

Multimodal perception⁶

Vision and touch are not separate resources for collecting different modalities of information. When the owner of the consignment shop in Excerpt 2 is doing inspecting the conditions of the tags for rewriting the prices on them, she is perceiving the tiny spaces as the sites for the prospective action of rewriting numbers rather than collecting information on the objects’ size, shape, color, and texture. In other words, she perceives the spaces as barely affording the prospective action, that is, as something more than the simple sum of elementary visual and tactile features. This is not only the analyst’s speculation; the designer also sees this perception of the affordance by the owner because otherwise, she (the designer) might not understand the owner’s inquiry as inquiring about the possibility of rewriting.

What constitutes the owner’s inspecting does not only involve what happens at her fingertips and in her gaze direction; it also involves the movements of the owner’s hand, head, and upper body. Perceiving the tininess of the tiny spaces on the tags is only

organized in the spatial and temporal voluminosity of the bodily configuration in which she also senses the movements and positions of her body parts relative to each other (Figure 4). The same is true of Mika's bodily configuration in Excerpt 3; Mika's seeing of the stationary front surface of the plate is organized in the spatial and temporal voluminosity of the bodily configuration involving Mika's proprioception of their body parts as well as her tactile perception of the plate that she is holding (Figure 7). Natsu's understanding of Mika's interest display is possible only in this bodily configuration (see Streeck, 2013, for the importance of kinesthesia in interaction).



Figure 4



Figure 7

As indicated with respect to Excerpt 2, the entire relevant bodily configuration also involves the recipient's body; in the bodily configuration, the owner is perceiving that the designer sees her perceiving. In Excerpt 3, Natsu turns the plate's front surface to Mika when Mika extends her right hand toward it. When Mika touches it, Mika's inspecting is not only seen by Natsu; their bodies are also connected through the plate that they touch simultaneously. Specifically, Mika can feel that Natsu feels that Mika is keeping the plate stationary in the air; in other words, Mika's inspecting the stationary front surface of the plate in front of her eyes is organized in the bodily configuration involving both bodies. Note, however, that the designer (Excerpt 2) is gazing at the object being inspected instead of at the inspector's' body. The bodily configuration is a

system of different orientations exhibited on different body parts, organized such that the orientations are convergent on one specific object at each time. In such a bodily configuration, the participants perceive each other's integrated orientations to an object to build a specific action in a distinct activity.

Perceiver's perspective

Nigro and Neisser (1983) observed that episodic memory (memory of an event that one actually experienced) may be formed either as a visual image viewed from the viewer's perspective (therefore, not including the viewer in it: "field memory") or as a visual image that includes the viewer in it ("observer memory"). Although they do not take a definitive position on which perspective the original perception takes,⁷ some philosophers lean toward asserting that the original visual perception only takes the perspective that does not include the viewer in it (field perspective). For example, Michaelian (2016, p. 137), full of insights in many respects, argues that "there may indeed be a greater divergence between the original experience and the retrieved memory in the case of observer memories," although field memory is also a product of reconstruction rather than a simple reproduction of the original experience. However, these arguments appear to assume that perception is mediated by some snapshot- or movie-like mental representation captured by the perceiver's fixed point of view or a camera eye (see Gibson, 1979, for the criticism of such ideas). As we indicated in the previous subsection, seeing is essentially multimodally organized within a bodily configuration with multiple bodies. The perceiver's perceptual experience includes experiencing that what they experience and how they experience it are perceived by other participants, and the perceiver perceives things precisely in the way that their

perception is perceived by others in order to accomplish a specific action with and for these others. In other words, perception in interaction essentially involves others' perspectives.⁸

The practice of doing inspecting the details of an object can be the practice of fixing the viewer's visual field or of obtaining a "snapshot" vision. The teacher in Excerpt 1 and the owner in Excerpt 2 are excluding the background from the visual field. In particular, Mika is fixing the plate in front of her eyes to create a stationary visual field. However, we have demonstrated that both the narrowing and fixation of a visual field are achieved in bodily configurations involving multiple bodies, in which a system of embodied orientations is convergent on a very limited visual field. In other words, obtaining a snapshot vision of an object is an accomplishment instead of the elementary perspective on which complex perception is built.

The unity of the body

How can one experience the unity of one's body despite the fact that different body parts have different experiences? This question is, as is evident now, an artifact of the empiricist assumption—namely, that our perception of the world is composed of information gathered by independent (five) senses. Waldenfels (2000) examined three approaches to the unity of the body. The first approach is empiricism, according to which the unity of the body is the association of different (elementary) senses. Empiricism poses the problem without providing a satisfactory solution to it. Second, according to Gestalt theory, the primary experience is grasping a Gestalt of stimuli rather than experiencing different stimuli with different senses. This approach would dissolve the problem. However, Gestalt theory is deficient in that it cannot explain

phantom limbs—the phenomenon of feeling pain (or other things) in the lost limb; the phenomenon lacks any stimuli on which a Gestalt can be built.⁹ Third, Waldenfels (2000) proposed his approach: The unity of the body is established in acting; a pianist who has lost a hand may still be oriented to the keyboard through the lost hand, and, as a result, the pianist may feel the lost hand. “The unity is provided by what is to be done at each time.... What matters is a practical unity that is built in doing and acting” (p. 115; translated from German). This third approach proposed by Waldenfels appears to be well supported by the empirical analysis we have presented; the unity of the body is an organizational feature of a bodily configuration oriented to a specific action, in which the participants perceive things in the world in an essentially multimodal manner.

Conclusion

This study has addressed a distinct phenomenon: doing inspecting the details of an object by leaning toward it and touching it. In the detailed analysis of two fragments in which a participant is doing inspecting an object that they have been seeing, we have demonstrated that seeing details does not necessarily aim at collecting detailed information but rather is a constitutive part of the ascribability of a specific action to the whole of the actor’s concurrent behavior. Furthermore, we have suggested that doing inspecting may be used as a disambiguating device on some occasions. This possibility serves as further support for the essential relationship between the ascribability of seeing to the viewer and the ascribability of a specific action to the whole configuration of the viewer’s concurrent behavior. We have also discussed some theoretical implications of our analysis, arguing that perception is essentially multimodal in a

bodily configuration involving multiple bodies; what one sees is interactionally organized in relation to the current and prospective actions, essentially involving the co-participant's perspective; and the unity of the body is provided for within the bodily configuration in which differently embodied orientations are appropriately distributed toward other bodies and the world in constructing a current action. Seeing details is not always relevant or favorable. Seeing details is seeing a physiognomy that an object shows in a specific action.

We have focused on a specific phenomenon—doing inspecting—but used it only as a lens through which to explore aspects of seeing. The potential limitation of this study is that we draw on only a limited number of examples; thus, we must leave a more thorough explication of the practice to subsequent investigations. However, we believe our claims are adequately grounded in the details of each case. Doing inspecting in the examined cases offered a perspicuous view of how ascribable perception is organized in a bodily configuration that involves multiple bodies and is oriented to a specific action.

Notes

¹ See Levinson (2012) for the notion of ascribing an action to an utterance (and/or embodied behavior).

² Certainly, the recipient of a request formatted as a “can-you” question may accept the request with an “I can” response. However, if it is an acceptance of the request, this response will sound like a hesitant rather than willing acceptance only focusing on the possibility or ability. In contrast, the designer in the example rather enthusiastically

responds with the emphatic adverb *zenzen*, translated as “absolutely.”

³ It is not very clear from the video what exactly the owner is touching; the shape of the owner’s fingers, combined with her talk, enables the conjecture that it is a tag.

⁴ Another support for this analysis is provided by the two appearances of *tatoeba* (“for instance”) in lines 07 and 10. It appears that the owner, in line 07, abandons the incipient indication of an example of changing prices that finally appears in lines 10 and 12. If the example appeared in line 10, it could not be heard as part of an inquiry about the possibility of rewriting prices in tiny spaces; it would rather be heard as more straightforwardly concerning changing prices as such. By letting the inquiry about the possibility of rewriting precede the indication of the example, the owner (re)constructs the example as the example of rewriting.

⁵ See Coulter (1979) for the notion of ascribing perception to a person.

⁶ Mondada (2021, p. 60) distinguished between multimodality and multisensoriality in the following way:

Multimodality refers to the multiplicity of linguistic and embodied resources that participants mobilize for interacting together in intelligible ways....

Multisensoriality refers to the sensorial experiences of participants as they engage in sensing the world and each other.

Certainly, when we read a book while holding it, we multisensorily (or multisensorially) experience the book. However, many “synesthetic” phenomena, such as seeing the rigidity of glass, may not be multisensory (although they are definitely multimodal) in that we do not tactilely feel the rigidity. Furthermore, all cases of seeing are essentially multimodal as far as they are always accompanied by the proprioception of the viewer’s movements and positions of eyeballs and other body parts. However, we are not sure

whether they always involve multiple senses. Therefore, we prefer the term *multimodality* as a general term for discussing the issues related to perception and perceptual experiences. Of course, this is not a criticism of Mondada’s use of the notion of multisensoriality.

⁷ “It is not clear how these experiences [e.g., experiencing one’s own reactions to defeat and failure from an observer perspective—a note added] are best interpreted—whether as a nonegocentric form of direct perception in Gibson’s (1979) sense or as the products of instantaneous reconstruction—*but it is clear that they exist*” (Nigro & Neisser, 1983, p. 469; emphasis added).

⁸ Moreover, what the owner in Excerpt 2 multimodally perceives in the temporal and spatial voluminosity of the bodily configuration oriented to the prospective action of rewriting in tiny spaces is the structures of the spaces that afford that action. From this, one may further argue that she “directly perceives” (Gibson, 1979) these action-affording structures independent of the images or representations that she may happen to have about them.

⁹ We would add that for the same reason, Gestalt theory cannot provide an adequate account of synesthetic experiences either. See also note 6.

References

- Auer, P. (2018). Gaze, addressee selection and turn-taking in three-party interaction. In G. Brône & B. Oben (Eds.), *Eye-tracking in interaction: Studies on the role of eye gaze in dialogue* (pp. 197–232). John Benjamins.
- Auer, P. (2021). Turn-allocation and gaze: A multimodal revision of the “current

- speaker-selects-next” rule of the turn-taking system of conversation analysis. *Discourse Studies*, 23(2), 117–140.
- Austin, J. L. (1962). *How to do things with words*. Oxford University Press.
- Böhme, G. (2013). *Atmosphäre: Essays zur neuen Ästhetik*. Suhrkamp.
- Burdelski, M., Tainio L., & Routarinne, S. (2020). Human-to-human touch in institutional settings: Introduction to the special issue. *Social Interaction. Video-Based Studies of Human Sociality*, 3(1), , 3(1). Retrieved from <https://tidsskrift.dk/socialinteraction/article/view/120247>
- Cekaite, A. & Mondada, L. (2020). Towards an interactional approach to touch in social encounters. In A. Cekaite & L. Mondada (Eds.), *Touch in social interaction: Touch, language, and body* (pp. 1–26). Routledge.
- Coulter, J. (1979). *The social construction of mind*. London: Macmillan.
- Coulter, Jeff. (1991). Logic: Ethnomethodology and the logic of language. In G. Button (Ed.), *Ethnomethodology and the human sciences* (pp. 20–50). Cambridge University Press.
- Drew, P. (2003). Precision and exaggeration in interaction. *American Sociological Review*, 68, 917–938.
- Gibson, J. J. (1966). *The senses considered as perceptual systems*. Houghton Mifflin.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton Mifflin.
- Goffman, E. (1979). Footing. *Semiotica*, 25(1-2), 1-30.
- Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers*. Academic Press.
- Goodwin, C. (1984). Notes on story structure and the organization of participation. In Atkinson, M. & Heritage, J. (Eds.) *Structures of social action* (pp. 225–246).

- Cambridge University Press.
- Goodwin, C. (1994). Professional vision. *American Anthropologist*, 96(3), 606–633.
- Goodwin, C. (1996). Transparent vision. In E. Ochs, E. A. Schegloff & S. A. Thompson (Eds.), *Interaction and grammar* (pp. 370–404). Cambridge University Press.
- Goodwin, C. (2007). Participation, stance, and affect in the organization of activities. *Discourse and Society*, 18(1), pp. 53-73.
- Goodwin, C. (2017). *Co-operative action*. Cambridge University Press.
- Goodwin, C. & Goodwin, M.-H. (1987). Concurrent operations on talk: Notes on the interactive organization of assessments. *IPrA Papers in Pragmatics*, 1, 1–55.
- Goodwin, C. & Goodwin, M.-H. (1996). Formulating planes: Seeing as a situated activity. In D. Middleton & Y. Engeström (Eds.), *Cognition and communication at work* (pp. 61–95). Cambridge University Press.
- Goodwin, C. & Goodwin, M.-H. (2004). Participation. In A. Duranti (Ed.), *A companion to linguistic anthropology* (pp. 222–244). Basil Blackwell.
- Goodwin, M.-H. (1997). By-play: Negotiating evaluation in story-telling. In G. R. Guy, J. Baugh, D. Schiffrin, & C. Feagin (Eds.), *Towards a social science of language: Papers in honor of William Labov* (pp. 77–102). John Benjamins.
- Heath, C. & Vom Lehn, D. (2004). Configuring reception: (Dis-)Regarding the ‘Spectator’ in museums and galleries. *Theory, Culture & Society*, 21(6), 43–65.
- Higashiyama, A. & Adachi, K. (2006). Perceived size and perceived distance of targets viewed from between legs: Evidence for proprioceptive theory. *Vision Research*, 46, 3961–3976.
- Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In G. H. Lerner (Ed.), *Conversation analysis* (pp. 13–23). John Benjamins.

- Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta psychologica*, 26, 22–63.
- Koiso, H., Tanaka, Y., Watanabe, R., Den, Y., 2016. A large scale corpus of everyday Japanese conversation: on methodology for recording naturally occurring Conversations. In: Proceedings of LREC 2016 Workshop on Casual Talk Among Humans and Machines, 9–12. Portoroz, Slovenia.
- Levinson, S. C. (2012). Action formation and ascription. In Stivers, T., & Sidnell, J. (Eds.). (2012). *The handbook of conversation analysis* (pp. 101–130). John Wiley & Sons.
- Lindwall, O. & Ekström, A. (2012). Instruction-in-interaction: A teaching and learning of a manual skill. *Human Studies*, 35(1), 27–49.
- Merleau-Ponty, M. (1960). Le philosophe et son ombre. *Signes* (pp. 201-228). Paris: Gallmard.
- Merleau-Ponty, M. (2012 [1945]). *Phenomenology of perception*. Translated by Donald A. Landes. New York: Routledge.
- Michaelian, K. (2016). *Mental time travel: Episodic memory and our knowledge of the personal past*. MIT Press.
- Mondada, L. (2018). The multimodal interactional organization of tasting: Practices of tasting cheese in gourmet shops. *Discourse Studies*, 20(6), 743–769.
- Mondada, L. (2019). Contemporary issues in conversation analysis: Embodiment and materiality, multimodality and multisensoriality in social interaction. *Journal of Pragmatics*, 145, 47–62.
- Mondada, L. (2021). *Sensing in social interaction: The taste for cheese in gourmet shops*. Cambridge University Press. The MTI press.

- Mortensen, K. & Wagner, J. (2019). Inspection sequences: Multisensorial inspections of unfamiliar objects. *Gesprächsforschung*, 20, 399–434.
- Nigro, G. & Neisser, U. (1983). Point of view in personal memories. *Cognitive psychology*, 15(4), 467–482.
- Nishizaka, A. (2000). Seeing what one sees: Perception, emotion and activity. *Mind, Culture, and Activity*, 7, 105–123.
- Nishizaka, A. (2007). Hand touching hand: Referential practice at a Japanese midwife house. *Human Studies*, 30(3), 199–217.
- Nishizaka, A. (2010). Self-initiated problem presentation in prenatal checkups: Its placement and construction. *Research on Language and Social Interaction*, 43(3), 283–313.
- Nishizaka, A. (2011). The embodied organization of a real-time fetus: The visible and the invisible in prenatal ultrasound examinations. *Social Studies of Science* 41(3), 309–336.
- Nishizaka, A. (2013). Distribution of visual orientations in prenatal ultrasound examinations: When the healthcare provider looks. *Journal of Pragmatics*, 51, 68–86.
- Nishizaka, A. (2014a). Instructed perception in prenatal ultrasound examination. *Discourse Studies*, 16(2), 217–246.
- Nishizaka, A. (2014b). Sustained orientation to one activity in multiactivity during prenatal ultrasound examinations. In P. Haddington, T. Keisanen, L. Mondada & M. Nevile (Eds.), *Multiactivity in social interaction: Beyond multitasking* (pp. 79–107). Jon Benjamins.
- Nishizaka, A. (2018). Aspect-seeing in the interactional organization of activities. In D.

- Favareau (Ed.), *Co-operative engagements in intertwined semiosis: Essays in honor of Charles Goodwin* (pp. 345–354). University of Tartu Press.
- Nishizaka, A. (2020a). Appearance and action: Sequential organization of instructions in Japanese calligraphy lessons. *Research on Language and Social Interaction*, 53(2), 295–323.
- Nishizaka, A. (2020b). Guided touch: The sequential organization of feeling a fetus in Japanese midwifery practice. In A. Cekaite & L. Mondada (Eds.), *Touch in social interaction: Touch, language, and body* (pp. 224–248). Routledge.
- Nishizaka, A. (2022). The granularity of seeing in interaction. *Journal of Pragmatics*, 190, 24–40.
- Rossano, F. (2012). Gaze behavior in face-to-face interaction. PhD Thesis, Radboud University Nijmegen, Nijmegen.
- Sacks, H. (1992). *Lectures on conversation*. Blackwell.
- Schegloff, E. A. (2007). *Sequence organization in interaction*. Cambridge University Press.
- Schutz, M. & Lipscomb, S. (2007). Hearing gestures, seeing music: Vision influences perceived tone duration. *Perception*, 36, 888–897.
- Searle, J. R. (1969). *Speech acts: an essay in the philosophy of language*. Cambridge University Press.
- Soto-Faraco, S., Kvasova, D., Biau, E., Nara, I., Ruzzoli, M., Moris-Fernandez, L., and Torralba, M., (2014). *Multisensory interactions in the real world*. Cambridge University Press.
- Stevanovic, M., & Peräkylä, A. (2012). Deontic authority in interaction: The right to announce, propose, and decide. *Research on Language & Social Interaction*,

45(3), 297–321.

Stivers, T., & Rossano, F. (2010). Mobilizing response. *Research on Language and Social Interaction*, 43(1), 3–31.

Streeck, J. (1996). How to do things with things. *Human Studies*, 19(4), 365–384.

Streeck, J. (2013). Interaction and the living body. *Journal of Pragmatics*, 46, 69–90.

Waldenfels, B. (2000). *Das leibliche Selbst: Vorlesungen zur Phaenomenologie des Leibes*. Suhrkamp.

Wittgenstein, L. (1953). *Philosophical investigations*. Blackwell.

Wittgenstein, L. (1980). *Remarks on the philosophy of psychology. Vol. 1*. Blackwell.

Appendix: Transcript Conventions

In all the excerpts, each line is composed of two or three tiers. First, there is a Romanized version of the original Japanese. Below this are phrase-by-phrase glosses where necessary. Finally, the third tier presents an approximate English translation. The first tier of the transcript utilizes Jefferson’s (2004) transcription system. In the second-tier glosses, the following abbreviations are used:

AUX auxiliary verbs

HOR honorific expression

P particle

POL polite

PST past tense marker

Some excerpts include annotations of the embodied conduct of each participant in the extra tiers designated by lowercase abbreviations, such as “stu.” The starting and ending

points of the movements are indicated by the sign |. Participants' gaze directions in the extra tiers are designated as ".g," e.g., "stu.g." In these extra tiers, a participant's or object's uppercase initial indicates that the gaze is directed toward this participant or object. Lowercase letters in these tiers indicate the transition of gaze directions.