Computer Supported Cooperative Work (CSCW) (2016) 25:235–278
DOI 10.1007/s10606-014-9215-0

From I-Awareness to We-Awareness in CSCW

Josh Tenenberg[1], Wolff-Michael Roth[2] & David Socha[3]
[1]*Institute of Technology, University of Washington Tacoma, 1900 Commerce St., Tacoma, WA 98402, USA (Phone: +1-253-692-5860; E-mail: jtenenbg@uw.edu);* [2]*University of Victoria, Victoria, Canada;* [3]*University of Washington Bothell, Bothell, WA, USA*

Abstract. Awareness is one of the central concepts in Computer Supported Cooperative Work, though it has often been used in several different senses. Recently, researchers have begun to provide a clearer conceptualization of awareness that provides concrete guidance for the structuring of empirical studies of awareness and the development of tools to support awareness. Such conceptions, however, do not take into account newer understandings of *shared* intentionality among cooperating actors that recently have been defined by philosophers and empirically investigated by psychologists and psycho-linguists. These newer conceptions highlight the common ground and socially recursive inference that underwrites cooperative behavior. And it is this inference that is often seamlessly carried out in collocated work, so easy to take for granted and hence overlook, that will require computer support if such work is to be partially automated or carried out at a distance. Ignoring the inferences required in achieving common ground may thus focus a researcher or designer on surface forms of "heeding" that miss the underlying processes of intention shared in and through activity that are critical for cooperation to succeed. Shared intentionality thus provides a basis for reconceptualizing awareness in CSCW research, building on and augmenting existing notions. In this paper, we provide a philosophically grounded conception of awareness based on shared intentionality, demonstrate how it accounts for behavior in an empirical study of two individuals in collocated, tightly-coupled work, and provide implications of this conception for the design of computational systems to support tightly-coupled collaborative work.

Key words: Awareness, Intentionality, Perceptual gestalt, Phenomenology, Shared intentionality, Socially recursive inference

1. Introduction

Awareness is one of the central concepts in Computer Supported Cooperative Work. What is it that a set of cooperating actors needs to be aware of in order to work together? After over 25 years of awareness research in CSCW (Gross 2013), there is an emerging view that (a) *conceptions* of awareness need to be grounded in philosophy so as to leverage prior conceptual work (Robertson 2002; Schmidt 2002b); (b) such conceptions need to be explored empirically, not only in trials of computational systems that provide awareness support for work at a distance (Calefato et al. 2012; Dourish and Belotti 1992), but in the everyday practices of collocated groups (Heath et al. 2002; Luff et al. 2008); and (c) such conceptions should provide insight in the construction of design principles and frameworks for technological awareness support (Gutwin et al. 2002).

Researchers have recently begun to clarify conceptions of awareness, helping the CSCW community address key questions, such as "[what are] the strategies competent cooperating actors employ to heed what colleagues are doing etc. How do they discriminate significant states, possible states, problematic states, etc.?" (Schmidt 2011 p. 35). In addition, they have helped to bring recognition to the fact that socially constituted forms of achieving awareness are specific to the practices, affordances, goals, or tasks of particular settings, accounting for why "designing generic awareness systems has proved fraught with difficulties" (Luff et al. 2008 p. 432). Yet what these conceptions share is a first-person perspective that black-boxes the intentionality of others, focusing only on the actions, communication, and resources that are "publicly available" (Robertson 2002), what we call *I-awareness*.

The problem with these first-person conceptions of awareness is that they do not take into account newer understandings of *shared* intentionality among cooperating actors that recently have been defined by philosophers and empirically investigated by psychologists and psycho-linguists. This newer conception highlights that individual intentionality is insufficient for the *socially recursive inference* that underwrites cooperative behavior. And it is this socially recursive inference that is often seamlessly carried out in collocated work, so easy to take for granted and hence overlook, that will require computer support if such work is to be partially automated or carried out at a distance. Ignoring the inferences required in achieving common ground may thus focus the researcher and designer on surface forms of "heeding" and attention but miss the underlying processes of intention shared in and through activity (Leont'ev, 1978) that are critical for cooperation to succeed. Shared intentionality thus provides a basis for reconceptualizing awareness in CSCW research, building on and augmenting existing notions of individual intentionality. And it is just such a reconceptualization of awareness, from "mutual awareness *of* something" carried out seamlessly and effortlessly (Schmidt 2011), to a "*shared* awareness of something that each recursively knows of the other" that we provide in this paper. Our key move is in going from first-person singular to first-person plural, from I-awareness to we-awareness.

The structure of our paper is as follows. First, we provide a brief summary of recent conceptions of awareness in CSCW research. We then provide a summary of philosophical and psychological research related to different forms of intentionality, with a focus on *shared* intentionality. We detail how socially recursive inference is at the heart of shared intentionality, and describe how human communication is underwritten by this socially recursive inference. With this grounding, we propose a notion of *we-awareness* based on shared intentionality. We provide a workplace study of collocated cooperative work in which the actors exploit socially recursive inference in the design and interpretation of the publicly available communication that they use for coordination. We identify not only their (individual) heeding strategies, but the socially recursive inference required for their tightly-coupled coordination. We then discuss how various forms of computational support for enabling this cooperative work to be carried out at a distance (non-collocated) embed

conceptions of awareness, and that a conception of awareness based on shared intentionality and socially recursive inference suggest new design possibilities.

2. Philosophical groundings of awareness in CSCW research

Researchers have commented on the multiplicity of uses of the term *awareness* in over 25 years of awareness research in CSCW (e.g. (Gross 2013; Robertson 2002; Schmidt 2002b)). Such multiplicity of uses has led to considerable conceptual confusion, a *babbling equilibrium* (Ostrom 2005). Heath et al. (2002 p. 318) comment "the difficulties in developing systems to support awareness do not simply derive from the limitations of technology, but rather from the ways in which we often characterize awareness and associated concepts such as mutual monitoring." In the introduction to the 2002 JCSCW special issue on Awareness in CSCW, after surveying the varied and different uses of the term *awareness*, Schmidt (2002b p. 287) remarks that "it is becoming increasingly clear that the term 'awareness' does not denote a set of related practices. In fact, it is hardly a concept any longer." Almost a decade later, he comments that awareness is "poorly understood" and "has barely been defined" (2011 p. 34). As a result, recent effort has been made toward *conceptual clarification* of this term. Luff et al. (2008 p. 408) note that empirical workplace studies, as much as they can provide specific design recommendations, are even more important to "contribute to a respecification of key concepts, like awareness, that are critical to an understanding of how technologies are used and deployed in everyday environments." And Robertson (2002) underscores the importance of making philosophical presuppositions explicit concerning such things as awareness, since such presuppositions underlie the choices and commitments embodied in the technologies that designers build to support cooperative work.

Schmidt has provided trenchant commentary on awareness in CSCW, explicitly linking the conceptualization that he proposes to its philosophical antecedents. "The first step towards some kind of conceptual clarity is to realize with the philosophers, from Husserl and Schutz to Wittgenstein and Ryle, that it does not make sense to conceive of 'awareness' as such, i.e., as a distinct (mental) entity. That is, the term 'awareness' is only meaningful if it refers to a person's awareness *of* something" (2002b p. 287). One of Schmidt's key insights is to note that the term "awareness" has been used in two distinct, and incompatible senses in prior research, thus leading to terminological confusion: as an "attention" concept, and as a "heed" concept. As an attention concept, "'being aware' is close to 'realizing', 'conscious of', 'noticing'" (2011 p. 34). Coupled with a cognitivist assumption of limited attentional resources, Schmidt argues that this construal of awareness makes it difficult to even recognize, let alone explicate, that there might be forms of human activity in which an actor can attend to one thing while at the same time monitor the relevant activities of others. On the other hand, as a heed concept, awareness is the mutual heeding that skilled actors perform in a material setting, "because they (normally) know the work and hence know what the others are doing, could be doing, should be doing, would

not be doing, and so on. They know the drill. Heeding what goes on is part of competent conduct" (2011 p. 35). It is this skillful, mutual heeding that Schmidt endorses, which then provides a basis for awareness research in CSCW. "Now, if this argumentation holds, this means that we, instead of searching for putative intermediate mental states, should try to identify the strategies competent cooperating actors employ to heed what colleagues are doing etc. How do they discriminate significant states, possible states, problematic states, etc.? What do they monitor *for* in the setting? What is ignored as irrelevant, what is taken into account?" (2011 p. 35) This conception, however, stops at the boundaries of skin and skull: one heeds what one can perceive of others without regard to their beliefs or goals.

The non-mentalist conception of awareness that Schmidt espouses can also be found in Robertson's discussion of awareness in CSCW (2002). Drawing explicitly on Merleau-Ponty's (2012) embodied phenomenology Robertson explains that "[f]or Merleau-Ponty perception is always an active, embodied process that is generative of meaning. . . . Perception always has a perspective. As an active process, it goes outwards into the world, from someone who is always somewhere at a specific point in time, taking hold of whatever is available in the environment that is *already* meaningful to that individual" (2002 p. 304). She explains that for Merleau-Ponty, perception is *skilled*, something learned through past experience in the world that in turn conditions subsequent perceptions of how situations "show up for us as requiring our response" (p305). Robertson explains that a person's materiality makes manifest his or her own activities, thus making these activities "publicly available" to the perception of others and the self. As embodied beings, people are made of the same "stuff" as the material world. As a result, people are simultaneously perceiving and perceived, what Merleau-Ponty calls *reversibility*. One consequence of reversibility that Robertson notes but does not further take up is that "any person being perceived can themselves perceive who is perceiving them" (p. 308), an issue that comes to prominence in the research study that we describe below. It is this simultaneous sensing and being sensed that leads to Robertson's characterization of the central dilemma of awareness for CSCW. "Reversibility does not hold in virtual space where the public availability of actions and artefacts does not rely on their being made of the same stuff but on their transformation and representation by the mediating technology" (p. 308). Mediating technology thus embeds the assumptions of its designers on just what it is that actors within a setting have to make available to one another in carrying out their cooperative work.

Workplace studies have provided considerable insight into the details of the "practices of awareness" (Button and Sharrock 2000; Goodwin and Goodwin 1996; Heath and Luff 2000; Heath et al. 2002; Luff et al. 2008). In a workplace study designed to explore the concept of awareness in CSCW, Luff et al. (2008) critique conceptions of awareness based on shared mental models, or the cognitive properties of individuals, as well as those that focus only on verbal communication. Doing so ignores the subtle ways in which "participants utilize what is public and visible in another's conduct to make sense of their activities, whether this is looking

at a public display, completing a computer keyboard command, writing part of a paper document, or noting the position of an object left on a desk or table" (p. 413). How the actors in a setting utilize these publicly visible resources is not a generic characteristic of technologies, nor is it "a way of being, a state, that is rather passive and peripheral" (p. 414). Rather, awareness concerns the particular activities in which the actors are engaged and the resources that the actors use and make available to one another within the setting (Heath et al. 2002).

The social is always implicated in Schmidt's discussion of awareness as a "heed" concept, Robertson's conception of the public availability of actions and artifacts, and Luff et al's characterization of the organized, meaningful social practices in which this heedful awareness is achieved. And yet in all of these philosophically grounded conceptions of awareness in CSCW, the minds of the individual actors are black-boxed. Heeding strategies, both the displayed and monitored, are the publicly available resources within a particular setting. Explicitly excluded are the beliefs and goals of self and other. Such mentalistic entities are simply *un*available to actors, and do not figure in the ways in which people are heedfully aware of one another in their ongoing, situated activities. The awareness of social actors is thus no more—and no less—than the aggregation of the awareness of the individuals who comprise the group in interaction, the aggregation of myriad atomic acts.[1] It is individuality of awareness, what we shorthand as *I-awareness*, that we wish to challenge and extend. Awareness, we claim, is a collective accomplishment, requiring and presupposing that actors take heed not only of what is external and public, but also of what is internal to one another

3. Shared intentionality and recursive social inference

Our conception of awareness as immanently social, what we call *we-awareness*, depends on recent conceptualizations of human intentionality that extend beyond the individual. These conceptions of shared intentionality, however, rely upon and extend earlier notions of individual intentionality, and we discuss each in turn.

One of the key concepts of the phenomenological philosophers is that of *intentionality*. Broader than the everyday sense of this concept—people *intend* to do things—this philosophical sense refers to *directedness* into the world. "The term originates from the Latin *intendere*, meaning 'to stretch forth' (Spinelli, 2000, p. 11). It speaks of our relationship with the world and how as conscious beings our experience is always *of* something" (Larkin et al. 2011 p. 323). Intentionality in this philosophical sense can be thought of as the meaningfulness, significance, or "aboutness" of human activity and thinking, including not only such things as goals, desires, and wants (intention in the everyday sense) but also plans, beliefs, and goal-

[1] We thank one of the anonymous reviewers for the phrase "the aggregation of myriad atomic acts" as a gloss on what we had been trying to express.

directed actions. "Intentionality, so defined, has no special connection with intending. Intending, for example, to go to the movies is just one kind of intentionality among others" (Searle 1995 p. 7).

Several commentators have placed intentionality at the heart of the phenomenological philosophy of Husserl (Dreyfus 1991; Larkin et al. 2011). According to Dreyfus (1991), Husserl's intentionality privileges individual mental content that serves to *represent* the world such that "this mental content gives intelligibility to everything people encounter" (1991 p. 2). This representational approach to intentionality continues to be pursued in such philosophers as Searle, who defines intentionality as "the capacity of the mind to represent objects and states of affairs in the world other than itself" (Searle 1995 pp. 6–7)

By contrast to this mentalistic approach, Heidegger, in *Being and Time* (1962), "countered that there was a more basic form of intentionality than that of a self-sufficient individual subject directed at the world by means of its mental content. At the foundation of Heidegger's new approach is a phenomenology of 'mindless' everyday coping skills as the basis of all intelligibility" (Dreyfus 1991 p. 6). Intentionality is inherent in and characteristic of all societal activities even though the individual may not be, and generally is not, consciously aware of it (Holzkamp 1983; Leont'ev 1978). Thus, in the actions of an assembly line worker who drives the same screws into frame after frame, intentionality in the form of the societal motive of car production is built in even though the worker may only think of driving screws. But because the societal activity and the actions that realize it constitute each other, every human action also has this societal intentionality built in. Such non-representational intentionality as a basis for human action is also an important aspect of Merleau-Ponty's philosophy, in which human embodiment makes possible this skilled coping (Merleau-Ponty 1962). "The life of consciousness—cognitive life, the life of desire or perceptual life—is subtended by an 'intentional arc' which projects round about us our past, our future, our human setting, our physical, ideological and moral situation. … In so far as I have hands, feet, a body, I sustain around me intentions which are not dependent upon my decisions and which affect my surroundings in a way which I do not choose" (Merleau-Ponty 2012 pp. 136, 440). Whether mentalist or non-mentalist, all of these phenomenological senses of intentionality can be taken as concerned with the individual, the first-person perspective.

Within the last few decades, a number of philosophers have defined notions of intentionality that extend beyond the individual (Bratman 1992; Gilbert 1989; Searle 1990). This is what Searle (1990) first described as "we-intentionality" (also called *shared cooperative activity* (Bratman 1992), *shared intentionality* (Tomasello et al. 2005), *joint intentionality* (Tomasello et al. 2005), and *collective intentionality* (Tomasello and Rakoczy 2003)), which is distinguished from but builds on the "I-intentionality" of Husserl, Heidegger, and Merleau-Ponty. "Shared intentionality … refers to collaborative interactions in which participants have a shared goal (shared commitment) and coordinated action roles for pursuing that shared goal" (Tomasello et al. 2005 p. 680). Searle argues that this shared intentionality "is a primitive

phenomenon that cannot be analyzed as just the summation of individual intentional behavior" (1990 p. 401). Searle makes clear that he does not appeal to any sense of "group mind" or similar. Searle argues that mutual belief or perception of a common goal (i.e. you and I each independently adopt the same goal) is insufficient for shared intentionality, as for instance when several individuals who are all enjoying an afternoon in the park all happen to move into a covered shelter *at the same time* when it begins to rain. Although their actions may appear coordinated, and each can form the belief that they themselves intend to enter the shelter and attribute similar intentions to others, it is not the case that the individuals have the intention to carry out their (individual) actions *together* as part of some larger enterprise. But the case is completely different, he points out, when there is an outdoor performance of a modern ballet in the same park, in which the troupe all converges in a covered shelter *as part of their coordinated activity in the performance*. Although these may be the exact same physical actions as in the first case, it is only in the second case that we would call this shared intentionality. For Leont'ev (1978), the collective activity, oriented towards a societal object/motive, determines the nature of an action, and, in the same move, provides it with a collective (shared) intentionality. Since the physical actions alone cannot distinguish between individual and shared intentionality, the difference is a result of what each actor has in mind.

Bratman (1992) makes explicit the conditions for shared intentionality, which Tomasello et al. (2005 p. 680) summarize as: "(1) the interactants are mutually responsive to one another, (2) there is a shared goal in the sense that each participant has the goal that we (in mutual knowledge) do X together, and (3) the participants coordinate their plans of action and intentions some way down the hierarchy—which requires that both participants understand both roles of the interaction (role *reversal*) and so can at least potentially help the other with his role if needed." One sense of *mutual* is "[o]f something that is an attribute of each of two or more parties independently: belong to each respectively" ("mutual, adj. and n." 2014). Tomasello (2014) argues that the requirement of shared intentionality goes further, what he calls *recursive social belief*: I have to have in mind not only that you and I have the same intention, that you and I will carry out our actions together (albeit in different roles), I also have to believe that *you* have such beliefs in mind about our activities. In short, "the goals and intentions of each interactant must include as content something of the goals and intentions of the other" (Tomasello et al. 2005 p. 680). Tomasello et al. (2012 p. 677) also use the term *recursive mind reading* to refer to this process, positing it as the basis for all cooperative activity. "Knowing together means engaging in some form of recursive mind reading (we each know that the other knows, etc.), which is the basic cognitive ability that enables humans to engage in all forms of joint and collective intentionality (Tomasello 2008, 2009), including joint attention, common conceptual ground, and all 'public' knowledge and activities."

Shared intentionality, and the socially recursive inference on which it depends, has a profound effect on the way in which actors shape and interpret their

communication. Grice (1975) was one of the first psycholinguists to recognize that speakers, in making their utterances, take account of the knowledge that they believe hearers already possess. This allows speakers to efficiently form their utterances (i.e. to rely upon hearers to "read between the lines") and provides hearers with a warrant for making inferences (what Grice called *implicatures*) from what is said so as to discern meaning far beyond the spoken word. Grice treated conversation as a fundamentally *cooperative* act, underwritten by cooperative norms embedded in speech, such as "be relevant" and "be perspicacious."

Given the back-and-forth of language, where speakers and hearers switch roles repeatedly, Clark and Brennan (1991 p. 135) extended Grice's cooperative maxims to a general principle governing *both* interlocutors in ongoing conversation, a *principle of least collaborative effort*: "In conversation, the participants try to minimize their collaborative effort—the work that both do from the initiation of each contribution to its mutual acceptance." One of the central tasks that speakers coordinate is *grounding*, i.e. the process of achieving common ground. Common ground is taken to be the recursive knowledge that each has of the other, i.e. "I know X" and "You know X" and "I know that you know X" and "You know that I know X," recursively all the way down.

From any speaker's perspective, common ground is an assumption but can never be guaranteed; I may shape my speech under assumptions about your knowledge that turn out to be false. To address this possible breakdown in communication, according to Clark, in their turn-taking behavior, speakers *present* an utterance, which they do not take as being mutually understood (i.e. common ground) until there is *acceptance* by the hearer. But this creates a problem: how does the recipient know that the presenter has understood her acceptance? "Note that the acceptance process is recursive. B's evidence in response to A's presentation is itself a presentation that needs to be accepted. But where does the recursion stop? . . . What keeps the process from spinning out indefinitely?" (Clark 1993 p. 154). Two processes prevent infinite recursion for practical purposes. First, the interlocutors "mutually believe that the partners [i.e. the contributing speaker and the hearer] have understood what the contributor meant *to a criterion sufficient for current purposes* [emphasis added]" (Clark et al. 1991 p. 129), i.e. the speaker and hearer settle for weaker evidence depending on context. Second, the interlocutors settle for decreasing strength of evidence as the conversation continues (Clark 1993).

Tomasello summarizes this recursive inferential process and the shared intentionality that it presupposes for both the speaker and the hearer. "I intend that you know something, and so I refer your attention or imagination to some situation (my referential act) in the hope that you will figure out what I intend you to know (my communicative intention). Then you, relying on our common ground (both personal and cultural), hypothesize abductively what my communicative intention might be, given that I want you to attend to this referential situation" (2014 p. 94). In doing so, Tomasello underscores not only the abductive process of socially recursive inference,

but also the underlying human motivations for shared intentionality. In addition, he couches his language in terms of communication *whatever the modality*, whether "the referential act" is carried out through gesture, speech, inscription, or otherwise.

To speak of such mentalistic activities as "inference" and "abduction" is not to assume that these are deliberate, conscious processes, or that they leave aside the tacit. Rather, it is to identify a particular kind of cognitive activity (perhaps short-circuited in common cases as Clark describes above) that takes place within people that underwrites their social activity. And, as far as current evidence indicates, no other species, even social ones such as the great apes, is able to engage in this socially recursive inference (Tomasello 2014).

In summary, people are both individually intentional, directed into a societally organized world that is already meaningful, and *collectively* intentional. They crucially rely upon the special, collectively intentional character of "Other Minds" (Ryle 1949) in entering into and carrying out cooperative activity. Under this view, awareness is not only about heeding and generating publicly available communicative actions and displays, it is also concerned with the intentions of collaborators and with carrying out the recursive social inference underwriting cooperative work.

In presenting the theoretical concept of shared intentionality, we do not mean to suggest that people are *either* individually *or* collectively intentional. As Vygotskij (2005) shows, individual intentionality, such as pointing, emerged in relations with others and thus *is* social. Collective intentionality and individual intentionality require each other, each constituting the other. There is considerable evidence emerging from recent studies in the primate and neo-natal labs that phylogenetically, human forms of shared intentionality emerged with the genus *homo* by building on great ape forms of individual intentionality (Tomasello 2014). And ontogenetically, shared intentionality arises at approximately one year of age in humans, building on the individual intentionality displayed by infants during their first year of life (Tomasello et al. 2005). Others have shown that the individual intentionality in humans *only* arises in patterned social interactions with other members of society, e.g. (Meshcheryakov 1979; Vygotskij 2005).

Among the research cited above, Clark and Brennan's concept of common ground has found mention in recent work in HCI and CSCW, e.g. (Convertino et al. 2011; Kirk et al. 2007), along with related notions such as "shared situational awareness" (Gilson et al. 1994), though not all of this is praiseworthy. For example, Koschmann and LeBaron (2003 p. 1) use a detailed analysis of the situated transactions and speech in a hospital operating room to conclude that "the notion of common ground represents a confusing metaphor rather than a useful explanatory mechanism." These authors base their critique on Clark's "contribution model," with its rigid definition of "presentation" and "acceptance" utterances that are belied by the sometimes ambiguous and overlapping speech that the authors observed in the operating room. Yet these authors use without question the term "mutual understanding," just as other authors in CSCW have discussed "mutual intelligibility" (Suchman 1987). Despite

the mentalistic implications of such terms as "understanding" and "intelligibility," for these and many other CSCW researchers, these notions stop at the boundaries of the perceivable, never intruding into the mental.

We suggest that the various forms of mutuality (of understanding, intelligibility, awareness) in the CSCW literature is a recognition, however tacit, of the socially recursive inference that underwrites them. For while this mutuality may be presupposed when people are collocated, it can break down more easily when people try to collaborate at a distance. The perspective of shared intentionality raises new concerns for both the technology designer and the empirical researcher. For the technology designer, the concern becomes how socially recursive inference can be supported technologically, particularly for individuals working at a distance. This is because taking into account the inference that underlies these heeding strategies increases the ability of designers to provide technologies to support this inference in situations that may lack the availability of the taken-for-granted heeding strategies that embodied copresence presupposes. For the empirical researcher, the operative questions concern the ways in which people in particular settings use and generate publicly available resources to contingently achieve common ground *and the socially recursive inference this achievement requires*, what we call *we-awareness*. The empirical study in the next section seeks to answer this question for one form of collaborative work.

4. An empirical study of we-awareness

Work in a variety of settings requires coordinating the effort of many people, often over long periods of time. Here, we examine different cases of work in the smallest group possible: a pair. We investigate two software developers engaged in *pair programming* in situ within a small software organization. Our point in choosing this particular setting is not for purposes of explicating or improving the software development process, but to provide a case for illustrating our concerns with shared intentionality and socially recursive cognition. To do so requires that we focus on the specifics of *some* setting. We also seek to characterize those general features of the setting that are characteristic of many forms of cooperative work.

4.1. Pair programming

Pair programming is a common practice in software development, where two software developers sit side-by-side, working together to program the same computer, sharing the same input and output streams (Beck 1999). This work is tightly coupled, in the same way that two people moving a table across a room (Schmidt 2011) is tightly coupled, with the actors being ongoingly engaged. In examining a dyad in tightly-coupled work, we remove some of the complexity of larger-scale coordinative work, such as the use of inscriptional and artifactual "coordination mechanisms" required by larger groups and longer time frames (Schmidt 2011). In

addition, in pairwise work, it is much more difficult for the participants to "free ride" on the effort of others to achieve the joint goal, something that is common in larger groups (Kopelman et al. 2002). Software developers already use a variety of computational systems to coordinate their work, including systems that might not have been specifically designed for supporting their cooperative work. These short, dyadic, computer-mediated cases, then, can be viewed as test-tubes in which we can examine, in the simplest way, the fine-grained structure of the communicative work that participants carry out in coordinating joint activity and the recursive social inference on which this work relies. And these observations may lead to specific insights about patterns of coordination associated with computational mediation.

Software development is a complex human activity. Because commercial software products are composed of up to millions of lines of source program code, and software developers produce on the order of tens of lines of source program code per day (Jones 1991), software development has long been recognized as requiring considerable social organization and coordination for its production (Cohn et al. 2009; Rooksby et al. 2009; Sharp and Venolia 2009). Most commonly, software developers organize their labor by subdividing programming tasks into small units of work that are programmed by individuals to be later integrated into successively larger program units. Although pair programming can be traced back at least 35 years, the practice of having *two* programmers working side by side to program a single unit of work has emerged only over the past 15 years as an important, though controversial, form of organizing software labor (Plonka 2012).

4.2. Study context

BeamCoffer,[2] the company where we collected the data reported in this study, is a 9 year-old software development company in the Seattle area of the United States whose main product serves over 13 million users worldwide. The company, owned by a non-US parent, has approximately 50 employees, about one-quarter of whom are software developers. The organization's product is a software system that helps friends and family share information. With its millions of users, the software product includes a significant backend Software-as-a-Service (SaaS) component, and has both a web-based version and client versions for Macintosh, Windows, iPhone, and iPad. In the discussion that follows, all terms first introduced in italics are those used by the participants themselves in this setting, while those introduced in quotation marks are labels that we have coined.

All employees work in a single large room without dividers. It is centered on the *developer stations*, a material configuration of tables, chairs, and computer equipment that allows four pairs of programmers to work in close proximity at the same time. Each programmer pair works at a *pairing station* having two keyboards and two mice controlling a single cursor and keyboard input on a single computer, and

[2] The organization name is anonymized, as are the employee names.

one or two computer monitors for output. When using two monitors, the monitors form a single, extended display. When pairing, developers sit side-by-side in front of the monitor(s) (Figure 1).

The software developers organize their pairing work in approximately two-hour contiguous time segments called a *pairing session*. The participants of a pairing session typically remain together throughout the entire session, but often switch partners with each new pairing session. There are three pairing sessions daily, and each is preceded by a *huddle* that lasts 2–5 min. The huddle is a meeting that typically includes all of the software developers and the program manager. Participants stand in a circle, and, in sequence, present what they have been working on, what they will work on next, and choose pairs for the next two-hour long pairing sessions. One of the features of this overall organization of software labor is that when a pair changes partners, the *task* that was being worked on by the pair, if not completed, will remain with one of the members of the pair, the *holdover*. The holdover will then be paired with a new partner (the *non-holdover*) who will work with the holdover for the duration of the pairing session. In addition, sometimes a pair is reconstituted in the middle of a pairing session, such as when one member leaves to carry out a more urgent task and another software developer takes this person's place. It is thus possible for the same task to be worked on by completely different individuals during the last pairing session of the day than the first. The holdover serves as a form of institutional memory concerning the immediately preceding state of the task and the work that has been done relative to it.

The data sources for this study include 360 hours of video of pairing stations, 20 hours of video from the huddle area, and four hours of video of other meetings. In

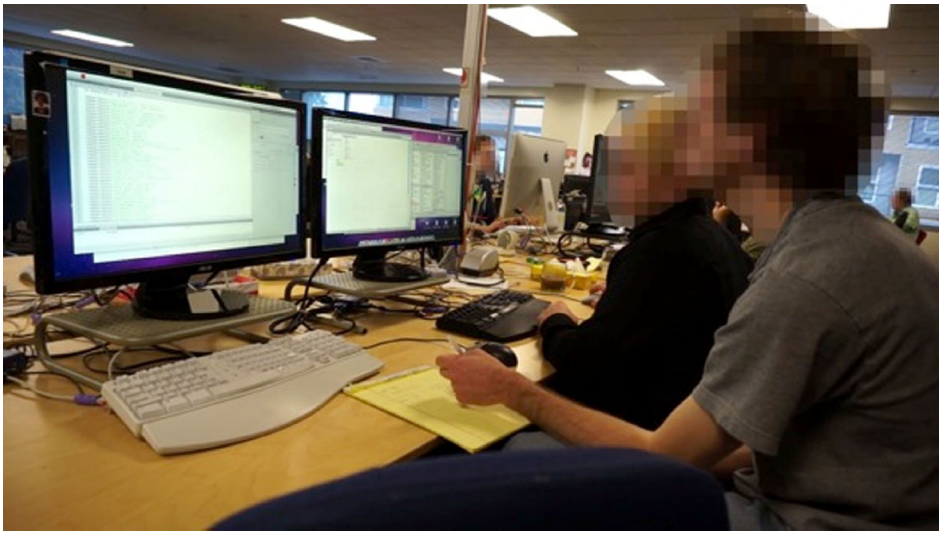


Figure 1. A pairing station at BeamCoffer

addition, the third author, a software developer with over 20 years of experience, spent many hours of ethnographic observation in the organization over the course of 21 visits from October 2012 through March 2014. In the process he took field notes, made several hundred photographs, and had a number of ethnographic interviews with key informants. Our primary method of analysis is *interaction analysis*, "an interdisciplinary method for the empirical investigation of the interaction of human beings with each other and with objects in their environment. . . . Its roots lie in ethnography (especially participant observation), sociolinguistics, ethnomethodology, conversation analysis, kinesics, proxemics, and ethology" (Jordan and Henderson 1995 p. 39). We focus on the transactional[3] work that a pair carries out in their joint activity, paying particular attention to the *semiotic resources* that they employ, "e.g. a range of structurally different kinds of sign phenomena in both the stream of speech and the body, graphic and socially sedimented structure in the surround, sequential organization, encompassing activity systems, etc." (Goodwin 2000 p. 1490).

4.3. Analysis

To exhibit the semiotic resources that pairs generate and rely upon, and the common ground that is presupposed and created in their transactional work, we analyze here three fragments from a 15 min period of the daily work at BeamCoffer. Figure 2 graphically shows the temporal relationship between these fragments within a larger episode of coupled work. We define a *fragment* as a contiguous temporal segment of transactional activity whose beginning and ending are researcher-determined. We provide transcripts of each fragment typographically set-off from the main text, along with analysis of these fragments.

Section 4.3.1 introduces fragment 1, in which a pair is just beginning a pairing session. In beginning their work together, a pair cannot presuppose that they are already in alignment. As Schmidt (2002a p. 29) makes clear, alignment itself is an achievement: "the practices of mutual alignment are not strictly speaking *effortless*. Actors' mutual alignment is predicated on selective and active monitoring and displaying." Precisely because the non-holdover does not know the present state of the work and the software, the pair must make visible various aspects of their cooperative work that must become common ground for the actions to be coordinated: Are we jointly committed? What are we doing? What is each of our roles? When do we carry out various actions? In this regard, coming into alignment is what has been called *articulation work*: "Articulation work is work to make work work. Or to be exact, articulation work is cooperative work to make cooperative work work" (Schmidt 2002a p. 462). Once alignment has been achieved, the pair can shape their activities so as to minimize communicative effort and thereby achieve the

[3] We use the term *transaction* to denote the relationship between people at work. This recognizes that relations and conversations cannot be reduced to the independent but interacting contributions of individuals but instead mutually implicate each other (Roth and Jornet 2013). This approach is consistent with the analytic stance described below of taking turn pairs as the minimum analytic unit.

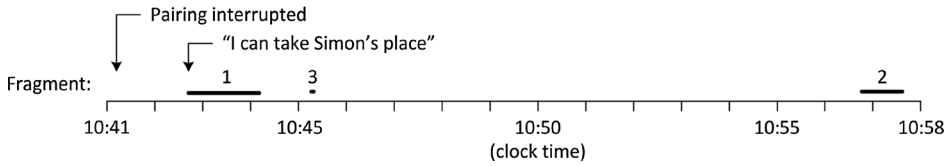


Figure 2. The location and duration of the three fragments discussed in the analysis

benefits that common ground affords, illustrated in fragment 2 (Section 4.3.2). In addition, the pair must maintain this alignment on an ongoing basis, repairing their activity when they detect they are out of alignment, illustrated in fragment 3 (Section 4.3.3).

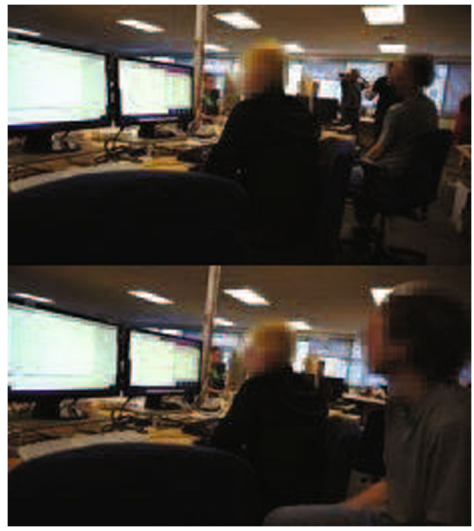
4.3.1. *Coming into alignment*

For you and me to work together, we have to align our significations, our intentions, so that our actions are mutually intelligible. We have to know at least what Clark (2005) suggests in carrying out joint activity: our joint goal, roles, and when and where particular actions will take place. But how does this joint activity get started in the first place, when we have not yet established a we-intention let alone the content of our joint goal and roles, nor the time and place of activity? Stated coarsely, our answer is that actors start from whatever common ground they already have (from prior experience, or a common cultural background), and use explicit forms of communication and mutual attention to build additional common ground. It is to such an instance of *coming into alignment* that we now turn, for it is here that the actors have to make visible and audible what they can (silently, seamlessly, effortlessly) presuppose in the rest of their joint activity.

Offering | accepting to pair. The fragment that we analyze[4] (fragment 1 in Figure 2) takes place shortly after the first huddle of the day. Hank had earlier started this pairing session with Simon, who has been with BeamCoffer for four years. After

[4] We use the following notational conventions for the transcripts, standard in conversation analysis (see as well Appendix A of (Roth 2013)). Unless modified, all words are written with lowercase letters. A number in parentheses indicates the length of a pause in the speech in seconds, while a period inside parentheses indicates a hearable pause of less than 0.1 s. Descriptions in double parentheses are transcriber's comments. Colons indicate lengthening of a phoneme, about 0.1 s per colon. Square brackets in consecutive lines by different speakers indicate overlap of speech between these speakers. Speech within angle brackets preceded by "p" (or "pp") standing for *piano* (or *pianissimo*) indicates lower (or much lower) speech volume than normal, as in "<<pp>scavenger hunt>." Speech within angle brackets preceded by "len" (or "all") indicates *lento* (or *allegro*), i.e. slower (or faster) than normal speed. A word inside parentheses ending with "?" indicates difficulty in hearing the word on the recording and that the word in parentheses is the closest approximation. A question mark inside a parenthesis is a word that could not be approximated. Capital letters indicate speaker's emphasis using a change in speech volume. An equal sign at the end of a word indicates that there is no hearable pause prior to the next word uttered. Downward and upward arrows indicate the pitch jumping downward and upward. The punctuation marks ",?;." indicate movement of pitch (intonation) toward the end of an utterance: slightly and strongly upward, slightly and strongly downward, respectively.

approximately 15 min, Simon is called away to participate in an interview for a candidate employee. Just prior to Simon's leaving, Danny, who has been working alone at the adjacent pairing station, requests assistance from Hank and Simon on a bug assigned to Danny. Danny sees Simon leave, and can hear Hank say "abandon ship".

1.1	Danny:	((Danny, in grey shirt on right, slides to the left side of Hank's pairing station, starts to speak in mid-slide.)) I can take Simon's place I can read fonts to you	

In stating "I can take Simon's place," Danny makes an explicit verbal proposal to do what Simon had been doing, which he glosses as "reading fonts". In making an offer to take a specific role within an ongoing activity, Danny also makes a bid to pair with Hank, to enter into a we-intention. In making his offer, Danny builds on the common ground that he believes that he shares with Hank concerning the work that Hank had been carrying out with Simon. Since Danny had been working nearby, he was able to overhear the work that Hank and Simon were doing. And in Danny's placement nearby, he was visible to Hank. Danny's utterance ("I can take Simon's place") thus signals that he knows the role that Simon was carrying out, and in making this role explicit ("I can read fonts") he shares his interpretation of Simon's actions with respect to an implicit goal.

In proposing to "take Simon's place," Danny takes an agent-neutral stance on the action (Tomasello et al. 2012; Tomasello 2009) whereby he reflects his understanding of Simon's activities as being not specific to Simon, but fulfilling a social role that can be taken by other agents, a "bird's eye view" of action (Tomasello et al. 2005). Speech provides this kind of perspective-taking that is difficult if not impossible to express using other semiotic resources. Agent-neutrality extends the normative

commitment beyond the two individuals ("we are committed because we entered into this joint agreement") to generalized norms of the group as a whole ("we are committed because that is what anyone in this organization is required to do who enters into a pairing relationship"). "Any praise or blame for an individual in a particular role is offered in the context of the standard that everyone mutually knows should be met. Thus, social practices in which 'we' act together interdependently in interchangeable roles toward a joint goal generate, over time, mutual expectations leading to generalized, agent-neutral normative judgments" (Tomasello 2009 pp. 91–2). The pair thus reproduces the social order that is normative within the organization, whereby most program code is generated in pair programming sessions to which both members of the pair are committed.

There is also non-linguistic coordinating work that occurs at the same time that functions along with the speech, and it effectively settles any concerns about *when* Danny will "take Simon's place." When Danny makes his proposal to pair, he is sliding his chair from one pairing station to another. His movement has semiotic dimensionality: it can be perceived at the same time that the linguistic offer is made. The movement indicates an embodied commitment to action, that is, Danny is not simply talking the talk, but walking the walk. In his physical placement, readiness to pair program with Hank *now* is apparent. In addition to the physical readiness that this placement affords, it configures the two of them *symbolically* as a recognizable pair, signaling to one another and the other software developers within the "horizon of observation" (Hutchins 1993) of others in the organization that Danny and Hank are engaged in a pairing session. Danny is thus literally and figuratively "taking Simon's place" in the action.

At the same time, Simon's "place" is in relation to Hank's place both physically and in terms of the intention of the programming task. While pairing with Simon and during Danny's move, Hank has been working in the right-hand position of a two-person pairing station. Hank is oriented to the right-hand monitor, and is using the mouse and keyboard to provide input to the computer. Danny saw and overheard Hank and Simon pairing together and subsequently saw Hank working alone at the pairing station. "Reading fonts," whatever that entails, is sensible only with respect to the work that Hank is doing within the larger task and project in which this action is embedded. In his utterance and body placement, then, Danny proposes not only his own role, but Hank's role as well, i.e. to continue doing what Hank was doing with Simon, in the "place" in which Hank has been doing it. In placing himself within the work, Danny also locates Hank's actions.

Establishing a work plan. As just described, in the first two seconds of this fragment, with both speech and body, Danny proposes to participate in an ongoing plan, proposes a division of labor in an agent-neutral fashion between the members of the pair, bodily commits to the activity in full view of his partner and organization,

and places himself in position to act. In the next part of this fragment, the details of the plan become an explicit issue of discussion.

1.1	Danny:	I can take Simon's place I can read fonts to you
1.2	Hank:	((Orients to right display.)) essentially it is a (1.99) ((Hank continues to work using the mouse)) a: (1.63) A:: scavenger hunt
1.3	Danny:	((Orients to right display))<<pp>scavenger hunt>
1.4		(4.03)
1.5	Hank:	I (1.66) don't see strange (lab man?)=
1.6	Danny:	so are we picking and choosing things, as opposed to the open everything (0.61) approach
1.7	Hank:	uh (1.13)<<len>picking>oh no notatall just we installed everything we could find
1.8	Danny:	okay

In turn 1.2, Hank does not respond directly to Danny's offer with acceptance or rejection. Rather, by starting to describe the work at hand, Hank indirectly and implicitly accepts Danny's proposal, thereby also signaling his commitment to the we-intention that Danny has already committed to. Hank states, "it is a scavenger hunt." "It" indexes the shared context that he and Danny are mutually oriented to on the computer display, the actions that Hank is taking with the mouse and his prior work on this task with Simon. This indexical expression presupposes shared common ground concerning the target of this reference. Hank is thus explicating the significance of the task at hand so that his own actions and what is being displayed on the computer monitor are intelligible to Danny. In addition, this signification of the larger goal will make intelligible Danny's actions in the role that he takes.

Danny repeats "scavenger hunt," the term that Hank used to describe the work, and then gazes at the computer displays for several seconds. Hank remarks on something that is "strange," thereby making publicly available something that is troublesome to him about the work. It is at this point that Danny reformulates what Hank has glossed as "scavenger hunt," this time not as an impersonal event that is occurring ("it is"), but as a question about what "we" are doing. Almost 10 s have elapsed since Danny made his initial offer to Hank, and most of that time has been spent with the pair silently gazing at the monitor, just as in the case described below, with its long silence.

But in this case, rather than having alignment, common ground about what they are doing, in turn 1.6, Danny indicates that Hank's actions are unintelligible, that in fact, Danny does not even know the goal that they are trying to achieve. For Danny and Hank to work *together*, the work of each must be made intelligible to the other, requiring that the activities of each must be congruent with the goal that they are trying to achieve and their jointly developed plan of action. Without this, Danny cannot integrate his activities with Hank's, the two "I's" cannot form a "we." In addition, in specifically using speech to form question (turn 1.6) and answer (turn 1.7), Danny and Hank are choosing a characterization for answering "what are we

doing?": is it a picking and choosing? An opening? An installing? It is this characterization, expressible in natural language, that provides the basis around which all of their actions will be intelligible. In his question "so are we …?" Danny is seeking common ground, and in so doing relies upon the common ground that he believes he already has with Hank. For Hank to eliminate one of the choices that Danny provides ("picking, oh no, not at all"), Hank has to believe that he understands (to the grounding criterion) what Danny refers to, how Hank's own actions might possibly be interpreted as "picking and choosing." And in offering "we installed everything we could find," he has to believe that Danny will abduce how this description glosses their joint activity, reformulating the two differently-stated alternatives that Danny has just proffered. In stating "Okay," Danny indicates that he has understood (to the grounding criterion) what Hank has just described as an account for "what we are doing."

Elaborating an account of past work. At this point, 31 s into fragment 1, Hank and Danny have achieved the following coordinating work: they have made a joint commitment to work together, a division of labor has been proposed, Danny has formulated two possible plan choices, Hank has eliminated one, and Hank has recounted past actions that provide continuity with larger sequences of activity and help to make the current activity intelligible. The semiotic resources used include speech (for formulating the specifics of plans, for specifying a division of labor in an agent-neutral fashion, for recounting the past) and mutually perceived bodily actions, both the placement of their bodies with respect to one another and to the material resources in the setting (for showing commitment to a proposed joint activity and role). Speech, perception, and embodied action co-occur and index one another.

At this point Hank continues to articulate what he and Simon had been doing:

1.7	Hank:	uh (1.13)<<len>picking>, oh no notatall just we installed everything we could find
1.8	Danny:	okay ((Danny orients to left monitor))
1.9	Hank:	and we ran it again and we had some (0.32) thirty ((orients to Danny's face)) (0.71) failures ((orients to left monitor))
1.10	Danny:	okay
1.11	Hank:	something like that (0.43) and so ((extends left arm horizontally from shoulder and points to left display with index finger)) we went through the list, TAGged the ones that ((drops hand to table)) dont match anything um sometimes its just a simple like spelling thing like here Andrew's script is really=
1.12	Danny:	bad
1.13	Hank:	bad ((tilts his head and flicks his left hand))
1.14	Danny:	((nods his head)) yeah awesome

In the utterance in turn 1.9, for the first time since the start of this fragment, Hank turns to gaze directly toward Danny, who does not turn to meet Hank's gaze but continues to orient to the left monitor. Hank notes that there were "some thirty

failures," and we might conjecture that this could become a joint problem for Hank and Danny to solve (i.e. to account for and/or eliminate these failures). It also answers, at the same time, the question why Hank is recounting these specific past actions: we tried to do this, and ended up with failures, which we now have to fix. In turn 1.10, Danny signals that he is following what Hank is saying.

Hank is in a physical position to see that Danny is still oriented to the left monitor, and in turn 1.11 takes advantage of Danny's orientation by extending his arm into Danny's field of view and pointing to the left monitor, where "the list" is displayed. Hank explains that there is a matching process that he and Simon were working on. What Danny can see is a list of font names within a script, but for this script to execute without error, each font name in the script must have the same name as one of the fonts in the filesystem, otherwise they do not "match". Hank explains that sometimes the reason for a mismatch is that the name in "the list" is a misspelling of the font name, which is "simple," suggesting that this can easily be fixed, e.g. by correctly spelling the font name in the script. Hank starts to refer to the script ("Andrew's script") as if to give an assessment about it and Danny completes Hank's statement with an assessment about the script ("bad"), which Hank immediately reiterates. When Hank reiterates this assessment, we can see that these two comments serve as an assessment | confirmation pair,[5] in that Danny offers the comment, which is confirmed by Hank.[6]

This assessment | confirmation pair is one of a set of such pairs whose use makes visible normative obligations that speakers have to one another. Prototypical pairs such as assessment | confirmation, offer | acceptance, and greeting | response are termed "adjacency pairs" in the conversation-analytic tradition (Sacks 1992). The "central characteristic [of adjacency pairs] is the rule that a current action (a 'first pair part' such as a greeting or question) requires the production of a reciprocal action (a 'second pair part') at the first possible opportunity after the completion of the first" (Goodwin and Heritage 1990 p. 287). Such pairs therefore reflect reciprocal obligations that interlocutors have toward one another in making their speech and activity intelligible to one another. In addition, in being sequentially organized, the first part of a pair serves as the immediate context for the second (see note 6). Because of this normative obligation, if the second part of a pair is missing, or delayed for too long, it creates a social breach and therefore cause for repair (Goodwin and Heritage 1990). In this fragment we see that the assessment | confirmation pair is done quickly and does not open a need for repair, as evidenced by Danny's "yeah awesome."

[5] We follow the suggestion to write irreducible analytic pairs using the Sheffer stroke "|" to indicate that each part of the pair co-implicates and determines the other part of the pair and, in this, the pair as a whole (e.g. (Roth 2013)).

[6] In a strong sense, the statement does not just belong to Danny, whose vocal organs have produced the sound-words, but also belongs to Hank, in whose ears the sound-words resonate at the same time (Roth 2014a, b). Hank's verbal articulation not only *implies* whatever he has heard, which has come from Danny, but also is *for* Danny. An articulation, therefore, cannot be ascribed to an individual but inherently belongs to both speaker and recipient. Speaker and recipient, thus, are oriented to and own, the same sound-words; this co-ownership constitutes, in part, the we-intention.

Marking coming into alignment. Adjacency pairs centered on assessments in particular have been well-studied, e.g. (Pomerantz 1984). "Although assessments may be seen as products of participation in social activities, the proffering of them is part and parcel of participating in such activities" (p. 58). Assessments, then, are *forms* of participation, not simply their outcomes. Their function in *this* transaction between Danny and Hank is not simply the sharing of experience, but the calibration of whether the pair, jointly working, is arriving on the same page. When Danny finds that his assessment accords with Hank's, Danny says "yeah awesome," which is sensible as a comment concerning the fact that he and Hank are currently aligned. This indicates to both that they are coming into alignment, though it is not yet clear if they have a plan of action, or if this will require additional work.

This fragment continues with Hank verbalizing "ankoom light," the name of a font:

1.12	Danny:	bad
1.13	Hank:	bad
1.14	Danny:	yeah awesome
1.15	Hank:	ankoom light (.) do not have it (.) unless Im crazy uh:
1.16	Danny:	you attempted to do a search and it didnt come up with anything
1.17		(4.66) ((Hank types at the keyboard. Both he and Danny orient to the right monitor as it displays the result of the actions that Hank has initiated.))
1.18	Danny:	wow=
1.19	Hank:	yes thats (.) what happens when you have 3000 fonts on the machine
1.20	Danny:	now now its only 781
1.21		(1.23)
1.22	Hank:	it didnt (.) it didnt install (all?) the t-t-fs or something like that (.) um:=
1.23	Danny:	okay

Since Hank and Danny are still oriented to "the list" of font names on the left monitor, when Hank says "do not have it," this is with respect to his earlier comment that some of the font names do not "match" a font stored in the filesystem. Danny then says "You attempted to do a search and it didn't come up with anything" (turn 1.16). Although this has the grammatical form of an assertion about what already transpired, it could also serve as a request to Hank to further elaborate his recount of what he and Simon already did, or a question for a simple confirmation, among other possibilities. The function that it serves in the transaction is not determined until after Hank's response of typing at the keyboard (turn 1.17), which indicates that Hank takes this as a request or directive to perform the action ("do a search") that Danny specified (Roth 2014). "Doing a search" will not (except in some inconsequential sense) change the state of the system, but provides information that Danny (and perhaps Hank) needs to determine how to proceed. It clearly shows the contingency of planning and acting, that making plans depends on learning things about the system and the world so as to determine how best to move forward.

Though Danny may have no idea if Hank had already done "a search" when he and Simon were working on this task together, in doing it now, Hank and Danny can see the results at the same time, and thereby take these results as common ground. In short, we have Danny's verbalized offer, and a silent but enacted response to which both mutually orient along with the results. We then have another assessment | confirmation pair, with Danny's "wow" (turn 1.18) being an assessment about the results of the action that Hank confirms with "yes" and an account for why this result occurs. In Hank's following comment that it "didn't install the t-t-f's," he elaborates his earlier account for why they saw the results they did, which Danny confirms with "okay." The sequential assessment | confirmation turn-pair followed by the account | confirmation turn-pair indicates (as it did a few moments earlier) that Danny and Hank are coming into alignment.

1.22	Hank:	it didnt (.) it didnt install (all?) the t-t-fs or something like that (.) um:=
1.23	Danny:	okay (2.54)
1.24	Hank:	s:o (1.66) I was of the opinion there were more places (0.52) um: at SOme point I guess we can go through
1.25	Danny:	((reaches for a yellow pad))
1.26	Hank:	and eliminate (.) ones that are spelling errors (.) figure out which ones are missing and sort of send out an all points bulletin who has these fonts on [their machine]
1.27	Danny:	[yeah lets do that]
1.28	Hank:	yeah ok (.) let's do that then (.) so (.) baybis (.) beebis

In turn 1.24, Hank starts with the hedge "I was of the opinion" that softens the assertion "there were more places." He does not explicitly define what is special about these places, but the continuation of this utterance in turn 1.26 indicates that these "places" are either the font names in the script that are either the result of spelling errors or do not name fonts on the filesystem. Hank then offers a possible plan that might be carried out"at some point": find all of the places where a font name in the script does not match a font on the filesystem (the ones that are missing), dispense with the ones that are spelling errors, and then try to locate the missing fonts with an "all points bulletin."

Starting joint programming. The "real work" of programming can start when there is sufficient evidence for participants that they are aligned and actually are in the position to collaborate. Even if there were no explicit verbal marker suggesting that pair programmers are aligned, a programming-related act implies that the condition for pair programming—being in alignment—has been achieved. In turn 1.25, Danny reaches for a yellow pad that is just to the left of the keyboard, pushes the keyboard closer to the monitor, sets the pad in front of himself where the keyboard had been, and holds a pen at the ready in his left hand (see Figures 3). Danny fully extends his body in reaching for the yellow pad, which is visible to Hank, who is orienting to the left monitor. Danny not only verbally agrees to Hank's proposal ("yeah let's do that"), but in so doing, offers this as their *current* plan to be enacted immediately, not

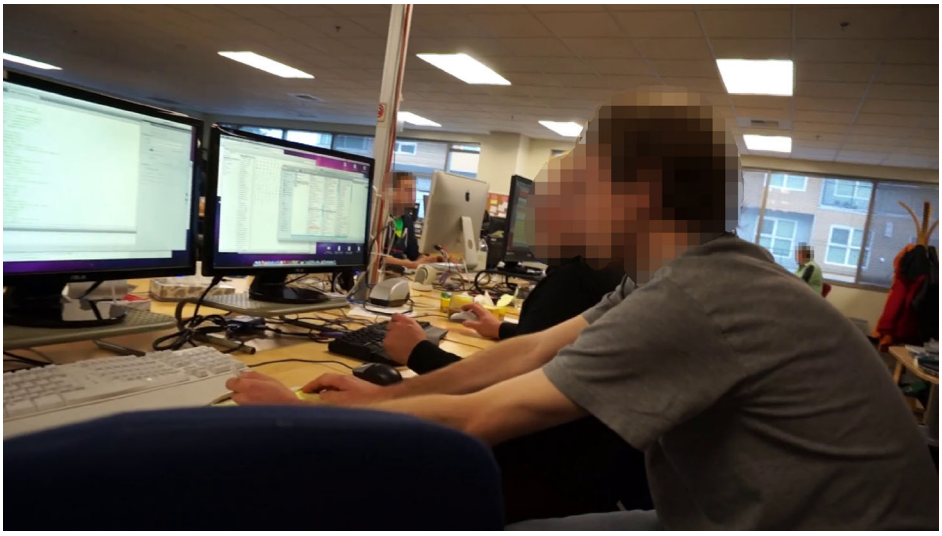


Figure 3. Danny ready to write on a yellow pad

at some hypothetical future time. In addition, by taking the yellow pad in hand, Danny is proposing (non-verbally) a division of labor. In turn 1.28, Hank agrees to carry out the plan immediately, and in stating "Baybis" starts to carry out the plan. At this point, 1 min and 29 s after the fragment begins, Hank and Danny quite apparently are in alignment, carrying out the work toward solving their joint problem that being in alignment enables.

Summary. Beginning to pair is a special kind of fragment, and it seems that to start the real work, there is considerable coordination that has to be done in advance. But because this coordination work is a condition for pair programming, it also is integral to it. Without the work of coming into alignment, little or no *pair* programming is possible. In coming into alignment, the participants have to make visible to one another what can be presupposed in much of the rest of their work. To summarize, joint commitments are made to work together, a plan of action is proposed and accepted, labor is divided. A variety of semiotic resources are marshaled, exploiting the co-presence of the pair. Though gesturing is little used, physical placement of the body is available to signal commitment, and readiness to engage in role-specific behaviors, e.g., hands placed on keyboard and mouse versus left hand holding a pen

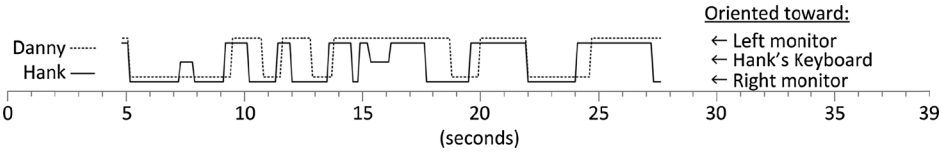


Figure 4. Danny and Hank's mutual monitoring

poised above a pad of paper. Speech is used in proposing the details of plans and to account for the results of actions taken on the computer. As the beginning of a pairing, the holdover recounts what has already been done to solve the current problem prior to the arrival of the non-holdover, thereby linking past, present, and future.

A considerable amount of semiotic work is achieved because the pair is mutually oriented to the computer monitors, and they can see (peripherally) one another oriented to the same monitor. There is, in essence, a recursive, mutual awareness that each has to the other and to the work. Co-presence affords this, and it is such a common feature of cooperative work that it is easy to take for granted. This recursive, mutual awareness also is substantially aided and evidenced by the adjacency pairs in their conversation and non-verbal communication. As we see in the analyses of the next two fragments, once the pair is on the same page, it needs to maintain the coordination to the work, and in so doing this mutual, recursive orientation continues to play an important communicative role.

4.3.2. *Being in alignment*

Coordination is not something that is established once and for all, where a plan is agreed upon prior to action and then executed as specified unproblematically (Schmidt 1997; Sharrock and Button 2011; Suchman 1987). The fragment here (fragment 2) occurs a few minutes after the one described in the prior section. Danny and Hank have already come into alignment, as detailed just above. Yet this state of alignment must be maintained: being in alignment is to be continuously *doing* alignment. Not only must the work itself be monitored, but the activity of the other person monitoring the activity must also be monitored, so that there is mutual awareness that can be indexically referenced afterward. In this fragment, Danny is to the left of Hank, in front of the left monitor and slightly behind Hank, and is holding a pen in his left hand. The left keyboard is pushed closer to the left display, and a yellow pad of paper is on the desk in front of Danny, who has been writing on the pad. Both Hank and Danny are oriented toward the leftmost monitor.

2.1	Hank:	so:=
2.2	Danny:	do you wanna just do a find and replace and we'll put some like (.) uh slash slash (?) the comment? so much like (?) ((orients directly at Hank, who remains oriented to the left monitor)) a marker there
2.3		(1.00)
2.4	Hank:	sure
2.5		((Danny orients to left monitor)) (1.18)
2.6		uh::
		((Hank types at the keyboard, and the results of the computation appear spread across both monitors))
2.7		(38.70)
2.8	Danny:	↑better-
2.9		(1.71)
2.10	Hank:	nice

In turn 2.1, Hank begins to speak, drawing out his "so" as he formulates his utterance. Danny immediately makes an offer of a specific plan to "do a find and replace." In his use of first-person plural ("we'll put some"), Danny establishes a we-intention. Although there is no rise in intonation at the end of this utterance, Hank treats this as an offer, which he verbally accepts in turn 2.4 and begins to carry out in turn 2.6. Danny and Hank are now mutually committed to this we-intention. Using the verbal modality, which each can hear, and which each can see the other has heard, establishes this commitment as a we-intention, i.e. it is socially recursive.

Using only the publicly available semiotic resources, can we as researchers be certain that Danny and Hank have grounded on their belief that they share this goal? Does Danny *really* believe that Hank is committed and that Hank believes that Danny is committed, and that Hank believes that Danny believes that Hank is committed, ad infinitum? We can no more do so than can Danny or Hank, as we (researchers and participants in the setting) are all constrained by what is publicly available. But, in beginning to execute the plan, and, as we will see, in aligning their perceptions, Danny and Hank presuppose common ground (Roth 2004). They thus act according to a grounding criterion that indicates that they are certain enough to continue.

During turns 2.6 and 2.7, Hank and Danny execute the plan. Almost 39 s transpire as they switch their gazes from one computer monitor to the other, with the output to these monitors changing dynamically. These 39 s are far from empty, as the events on the monitors to which they are jointly oriented constitute semiotic resources for maintaining and controlling alignment. Because the monitors are large and Danny and Hank sit close to them, they have to reorient their heads to turn their focus from one monitor to the other. Viewing Danny and Hank following the action from one monitor to the other is much like viewing spectators watching a tennis match as they follow the ball from one side of the net to the other Figure 4. Figure shows Hank and Danny's head orientations toward the left monitor, Hank's keyboard, or right monitor. The timeline starts at the beginning of the 38.7-s period of silence in turn 2.7. The solid line shows Hank's orientations and the dashed line shows Danny's. A line is at the top when the person represented is facing the left monitor, in the middle when he is facing his keyboard, and at the bottom when he is facing the right monitor. The dashed line shows the same for Danny's orientation, who looks only at the left and right monitors. The orientation lines begin at five seconds into the 38.7 period, the point at which Hank finishes his first set of keystrokes that initiate the computation and end at 27.2 s, when Hank and Danny no longer track one another's heads. The line orientations are plotted in intervals of 0.1 s, a breakpoint of time measurement in conversation analysis (i.e., $t<0.1$ s are indicated in the transcription as "(.)," whereas times $t>0.1$ s are measured and indicated explicitly in the form "(1.71)"). Each transition from one orientation to another starts when the person begins to orient his head or body to the new location.

During the first 5.1 s, and after 27.2 s, Hank and Danny do not track one another's head movements; instead, they move their heads independently of one another,

glancing every second or two at one of the computer displays. From 5.1 s through 27.2 s, however, Danny tracks Hank's gaze orientation: Danny changes his orientation 10 times, each time orienting to the same computer display as Hank and lagging Hank's change by approximately one-half second.

At 38 s into this silent period, one second before its end, the computation on the monitors stops and a list is displayed on the right monitor, to which both Hank and Danny immediately orient. One second later, Danny says "better" (turn 2.8). Danny's assessment is emphatic, sharply higher in pitch than his previous utterance (turn 2.2). In addition, Danny's assessment rises steadily in pitch, which can be heard as an offer to Hank for agreement. In turn 2.10, Hank confirms Danny's assessment with his comment "nice."

What inference is afforded by the alignment that Danny and Hank achieve throughout this silent period? Both are oriented toward one or the other monitor. Given that the human binocular visual field is approximately 180° in the horizontal dimension (90° laterally away from the nose for each eye) (Spector 1990), in orienting their heads to the monitors in front of them, each is able at the same time to see the orientation of their partner's head. But vision is not required to be aligned in activity: intonational, rhythmic, and other features available in the setting suffice to signal alignment and coordination, and perfect alignment of actions has been observed when there was no visual access (Roth and Tobin 2010; Roth 2011). This means that each operates on the assumption that the other is looking (more or less) at the same thing at the same time. The simultaneity is particularly important in *this* activity (i.e. pair programming), where events are changing quickly on the computer monitors, whereas such simultaneity may be less important in another activity, e.g. when object state is considerably more stable. If Danny and Hank believe that the other is sufficiently competent, they can also infer that both of them are not only looking at the same thing but are in fact *seeing* the same thing. Seeing, as Goodwin (1994 p. 606) points out, is a skilled activity, "a socially situated, historically constituted body of practices through which the objects of knowledge which animate the discourse of a profession are constructed and shaped." They are thus not simply "looking at the same monitor" but are rather both "seeing that no font mismatches have been found," (or whatever this computation happens to be about as experienced from their professionally situated vantage points).

The mutual alignment affords one additional inference, and this is the socially recursive one: that they are seeing the same thing, and that each knows the other knows this, ad infinitum (Roth 2004). For example, since Danny can see Hank oriented to the same monitor, Danny can operate on the assumption that Hank can also see that Danny is looking at the same thing, and can also operate on the assumption that Danny can see that Hank can see this, and so on. And similarly from Hank's point of view. This requires that Danny and Hank each treats the other not only as having similar physical capabilities, but that each operates as if the other can carry out socially recursive inference, i.e. that the other is someone who is collectively intentional.

Can we be certain that Danny and Hank have made these inferences? What most strongly evidences that they each operate based on the assumption of common ground sufficient to move their work along (i.e. to the grounding criterion) is in their verbalizing their assessments at the end of this fragment. In saying "better" in turn 2.8, Danny is not simply breaking the silence, not simply expressing his opinion about the state of affairs. Rather, he is saying something *for* Hank as part of their cooperative work. To make this inference process more explicit, we provide Tomasello's earlier quote (2014 p. 94), substituting the specifics of the case here to gloss what the "better" implies. "I (Danny) intend that you know that I believe that our work is producing a better outcome, and so I refer your attention by verbalizing 'better' out loud (my referential act) in the hope that you will figure out what I intend you to know (my communicative intention). Then you (Hank), relying on our common ground (both personal, i.e. that we have both looked at the same monitors at the same time, and cultural, i.e. that this allows us to *see* the same things at the same time), abduce that I (Danny) intend for you to understand my assessment of the events that we have just been seeing with respect to our joint goals, given that I want you to hear what I have just spoken, since otherwise I would have remained silent." And in making his assessment, Hank undergoes the same inference process but with the roles reversed, so that after both assessments, both can ground the belief that they are achieving their shared local goals.

How are we to interpret those times when Danny and Hank are not aligning their head orientations with one another, at the beginning and end of the 38.7 s silent period? What is important to note is that their non-alignment is also undertaken based on the assumption of common ground, since each is able to see (through an analogous argument to that above) that he is not seeing the same thing as the respective partner. But this seems to have no effect on their coming into alignment, at 5 and 38 s. This is because each is intermittently monitoring the work that is displayed on the computer monitors, to see if "anything is going on" that requires more active and ongoing monitoring.

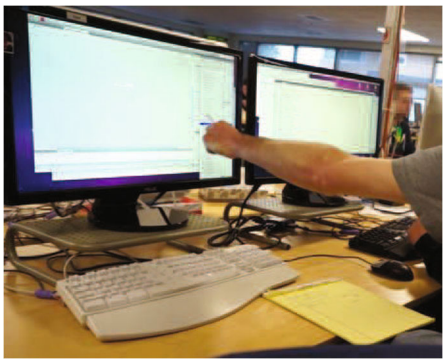
4.3.3. *Repairing alignment*

In maintaining alignment, there are times when repair work has to be carried out. In the case here, Hank and Danny have to "align" not only metaphorically, but literally, to bring their attention to the same line of text in preparation for operating on this text. Such precise alignment work is required because of ambiguities in gestural and spoken deixis, requiring fine-grained monitoring and repair. In this next fragment, we can see that achieving this alignment requires synchronized work using several interleaved forms of communication between the pair.

This fragment (fragment 3) occurred a few minutes prior to fragment 2 (see Figure 2) described in the previous subsection. Danny and Hank are both oriented to the left monitor. In a window on the left-hand side of the monitor is a list of font

names. Danny is holding a pen in his left hand, having just written with it on a yellow pad on the computer table in front of him.

3.1	Danny:	((Just before he starts talking, Danny moves left hand that is holding a pen so that the pen points to a specific item on a dropdown menu on the left monitor)) I bet you if
3.2		((at apex of point, taps pen on screen))
3.3		((Hank selects item on list that is four items below Danny's point, which is highlighted on the display))
3.4	Danny:	you (go?) ((starts to withdraw hand))
3.5	Hank:	((Hank uses mouse to move cursor two elements higher on the list))
3.6	Danny:	bidoni
3.7	Hank:	((Hank moves up two additional elements on list, stays there))
3.8	Danny:	m-t-m black



Before turn 3.1 of this fragment Danny and Hank had already agreed on their plan, but the specific item that they need to operate on is at issue. As their cooperative work suggests, this needs to be identified, agreed on, located, and selected. In his first turn, Danny's speech and gesture are synchronic. The gesture is deictic, and exploits Hank's orientation to the left display, in that Danny's entire body moves forward through Hank's field of view. Danny utters the first part of a counterfactual statement ("if you"), as if it were to be completed by some action and its effect: "if you do X then Y." But Danny does not complete the counterfactual, thereby calling into question that the purpose of this statement is to make a prediction about what will or might happen if a particular action is taken. Rather, Danny's pen-as-pointer "goes" to a location on the list of fonts on the display. Given Hank's response of moving his cursor to the neighborhood of the target location, Hank takes Danny's deictic gesture as completing Danny's own utterance, in that the gesture functions with the words to indicate "go here." That is, the utterance-gesture pair can be read as "go to the location that is the target of the pointing pen," since that is where the work to be done is located.

In close coordination with Danny's speech-gesture proposal, Hank "goes" with his cursor to an item close to where Danny's pointer came momentarily to rest. In making this movement and visibly highlighting an element near the pointer's target,

Hank accepts Danny's proposal without uttering a word. But Danny's and Hank's "here" are different, which Danny recognizes, and thus he initiates a repair. Danny uses the auditory modality to name a specific element on the list, "bidoni,", to which Hank quickly moves. Danny's utterance can now exploit Hank's visual fixation near where he has already directed the cursor, something that was not exploitable prior to Danny's deictic gesture and Hank's prior movement. In this, Hank's movements of the cursor represent a "placing-for" (Clark 2005), in that he moves the cursor to a specific location because this is a prerequisite for operating at that location. In the situation that Clark examines where a pair is collaborating to build a piece of furniture, this is done with material objects (screwdrivers, parts) that need to be placed prior to subsequent use. In the case here, coordination is in the virtual (rather than the material) world, so that the "placing-for" is performed with the cursor in selecting a piece of text in a document. What is important about a "placing-for" is that when visible to the other person it conveys semiotic content. "Placements acquire interpretations from the places to which the objects are moved" (p. 517). Placing-for follows a preparatory principle, in that "[t]he participants in a joint activity are to interpret acts of placement by considering them as direct preparation for the next steps in that activity" (Clark 2003 p. 260).

In summary, then, we can say that although Danny and Hank are already aligned at the start of this fragment, for this piece of work, they have to not only be metaphorically aligned, but literally on the same line (of text) as well. Danny uses physical gestures and speech that complement and complete one another to direct Hank to a specific location. Hank uses the mouse for placing the cursor preparatory to acting with it, which, in its visibility to Danny takes a role in the "conversation" that the two are having concerning the specific location of the next operation. They thus combine a variety of semiotic resources to give this fragment its orderly, sequential character (Sacks et al. 1974).

5. Discussion

In conceptualizing awareness as a "heed concept," Schmidt (2011 p. 35) provides a basis for how awareness research in CSCW can proceed. "[I]nstead of searching for putative intermediate mental states, [we] should try to identify the strategies competent cooperating actors employ to heed what colleagues are doing etc. How do they discriminate significant states, possible states, problematic states, etc.? What do they monitor *for* in the setting? What is ignored as irrelevant, what is taken into account?"

Our central claim, however, is that the heeding of publicly available actions from an individual perspective is insufficient to account for the achievement of some forms of cooperative activity. Rather, in pair programming (and we suspect similar kinds of activity) the pair often uses a heeding strategy that presupposes socially recursive inference, i.e. that each has to take account of the intentionality of the other. I-awareness, even when aggregated across multiple individuals, is insufficient to

account for the we-awareness that people develop and exploit in tightly coupled work. A key reason that awareness has been so central to CSCW is that tools to support cooperative activity embed designers' assumptions about the kinds of awareness that the tool users require. If tool builders do not take different forms of awareness into account in the tools that they build to support this work, then there may be breakdowns in tool-mediated work requiring costly communicative repair. In making our argument, we delineate the different forms of awareness (I-awareness, aggregated I-awareness, and we-awareness), and argue for and evidence breakdown in tools associated with each.

5.1. I-awareness

We take I-awareness to be as Schmidt describes above, concerned with "heeding strategies," with what is monitored, what is taken into account. Looking across the pairing sessions, except for rare moments, each member of the pair orients his body and head to one of the computer monitors on the table in front. When they speak to each other, they almost never look at one another as in everyday conversation, but keep their gaze oriented to one of the monitors. We can thus say that the most important heeding strategy is to *see the computer display and where the other person is oriented at the same time*.

Several software systems have recently been developed for supporting two pair programmers working at a geographic distance, what has come to be called *distributed pair programming* (Baheti et al. 2002; Dajda and Dobrowolski 2007; Flor 2006; Hanks 2008; Salinger et al. 2010; Schummer and Lukosch 2009; Stotts et al. 2004). One of the first such systems consists of commercial-off-the-shelf (COTS) desktop sharing software and real-time audio (Baheti et al. 2002). One member of the pair edits the software (whom we call the *editor*) while the other remotely views the changes that appear in real-time (whom we call the *remote viewer*). Given these tools, there appears to be sufficient support for the remote viewer to achieve the identified I-awareness heeding strategy (monitor the computer display and where the other person is looking), but only when the editor is actively changing what is displayed. Under this condition, the remote viewer can take the location of textual changes as a proxy for where the editor is looking. This condition can break down, however, if the editor moves quickly to another part of the display or stops her editing activity.

Unfortunately, there is insufficient reporting of the transactional work at a fine-grained level (e.g. with line-by-line transcripts) when programmers use distributed pair programming systems equipped in this fashion to evidence our theoretically-informed prediction. But we can get a sense of the breakdown that this kind of mediated work at a distance results in by looking at the transcripts that are reported when non-collocated collaborators use these same computational means (screen sharing and real-time audio) in carrying out a shared editing task. Dourish and Belotti provide such an example in describing the work of designers using their editing system *ShrEdit*, "a synchronous, multi-user text editor" (1992 p. 4). *ShrEdit* allows each user to create

shared windows that appear on all of the concurrent users' displays and in which each is able to edit. As with distributed pair programming system described above, there are no telepointers, but any changes that one person makes appear almost instantaneously in the shared screens of the others. Consider the following exchange that the authors report (p6), along with their commentary that precedes the transcript.

> There were also problems with informing others about what you were doing. Users often volunteered such information to the group, as below:
>
> Two designers are working together on part of the document whilst the third is attending to another part. The third designer alerts the other two to a change, as opposed to an addition, he wants to make.

4.1	D1:	Lets make the first, designer stamps from preset selection.
4.2	D2:	OK… Now I'll copy this; I'll cut this…
4.3	D1:	Yeah cut that stuff below and put it in phase three.
4.4	D2:	…I can't cut that, I'll just copy that down to…
4.5	D3:	I don't think there's 'no salaries to pay', it's 'fewer'. You've got to have some kind of fix it.
4.6	D1:	Huh?
4.7	D1:	What are you doing [D2]?
4.8	D2:	What? … I'm doing… I'm down in the fax stuff.

Although left unspecified in the original, we use the female gender to reference the different designers. In turns 4.1 and 4.2, we see the same kind of offer | acceptance pair for a plan of action as we saw between Danny and Hank in preparation for work. Turn 4.2, however, is unlike what we see in the collocated programming fragments analyzed above (as well as in the other pair programming episodes that we have examined), in the explicit verbalization of what D2 herself is doing, a point to which we return below. D1 and D2 continue in 4.3 and 4.4, using "that stuff" and "that" unproblematically to index particular locations at which they are operating. The use of indexicals indicates that the pair is operating from assumptions of common ground concerning where they are operating. D1's utterance in turn 4.6, signals a communicative breakdown, with the expression of "Huh?", followed in turn 4.7 by "What are you doing?" D2 responds first by stating "What?", which is followed by the start of an explicit verbalization of activity, "I'm doing." D2, however, does not complete this description, and instead provides an explicit description of *where* she is operating. D2 thus signals that she interprets D1 as misunderstanding not the intent of a viewed operation, but that no such viewing is even taking place. In stating "I'm down in the fax stuff" D2 treats this *as* a response to "what are your doing," i.e. were D1 to look in this location then D1 would see what D2 is doing. Such a response by D2 is intelligible given that both were working unproblematically until turn 4.4 when D2

says "I'll just copy that down to," suggesting that D2 then went "down in the fax stuff" to complete the copy operation. This suggests, then, that D1 "lost" the location of the operation when D2 moved from the source location of the copy to the target location for the paste.

We use this fragment from the reporting of *ShrEdit* not to critique a system that is over two decades old, but to illustrate the kind of breakdown that our theoretical argument predicts will arise in the use of pair programming systems that provide similar mediational means. This breakdown is as predicted given this form of electronic mediation, since one member of the pair loses track of where her partner is working.

5.2. Aggregated I-awareness

What we call aggregated I-awareness is simply to take account of the I-awareness of all of the cooperating actors, recognizing that they occur simultaneously. Although these need not be symmetric (i.e. different actors can employ different heeding strategies), in the pair programming case that we examined, the heeding strategy for each member of the pair is symmetric: Each person can at the same time see the computer monitor and where the other person is oriented. Although this is supported for the remote viewer with the distributed pair programming tool described above, the person who is editing has no visual indication of where the remote viewer is looking and hence cannot carry out the identified I-awareness heeding strategy.

In an attempt to reduce communication breakdown, software tools were subsequently developed for distributed pair programming that, in addition to the screen sharing and real-time audio of the system described just above, added *symmetric control* (i.e. either programmer could edit) and *telepointers* (Dajda and Dobrowolski 2007; Hanks 2008). Telepointers provide a form of "virtual embodiment" (Robertson 1997). "Telepointers are the simplest form of embodiment, and show the location of each team member's mouse cursor. Telepointers are effective at conveying awareness information, since the mouse cursor is the primary means by which people carry out actions in computational workspaces. In addition to simple cursor location, telepointers provide implicit information about presence, identity, activity, and even the specifics of an action" (Gutwin and Greenberg 2001 p. 13). In active use, a telepointer thus acts as a proxy for where the person who controls it is looking, under the assumption that this person looks at the location where she is operating.

Again, there is insufficient reporting in the literature of the breakdowns that occur when programmers mediate their work with this tool. What then, can we predict about the ability to carry out the identified aggregated heeding strategy and the likely communication breakdown?

When both programmers are using their telepointers in close proximity, they can carry out the identified, aggregated heeding strategy, each seeing and being seen at the same time. Yet the use of telepointers can break down when one programmer moves her activity to another part of the screen. Indexical references by this moving

actor might either not be voiced (in anticipation of high repair costs), or, if voiced, require repair as in the Dourish and Belotti fragment discussed above. In other words, telepointers can be "lost" to the vision of a remote partner.

There is an additional problem that our pair programming study points out: what inferences can each programmer make about the other when *neither* of them is actively using their pointer? There are many times when the computer displays behavior over time that is autonomous from either of the programmers, as illustrated in the second fragment analyzed above during the 38.7 s of silence The computer itself has perceivable activity in response to Hank's earlier keystrokes. Collocated programmers such as David and Henry can rely on their simultaneous observation of both the computer display and one another's head orientation to support the use of indexicals and to coordinate their activity. But these very cases, when the work self-signals, are unsupported by telepointers, screen sharing, and real-time audio.

5.2.1. *When work self-signals*

Practical action not only gets work done, that is, has ergotic function, it also has symbolic function, signaling, at a minimum, itself (Cadoz 1994; Roth 2003). In fragment 3 in which Danny and Hank are orienting to the same line, Danny points to a location on the monitor, to which Hank tries to orient with his mouse-controlled cursor. What this means is that Hank's *actions* when mutually perceived, take part in turn-taking behavior, supplanting speech or gestures for this purpose. We see something similar in the first fragment analyzed in Section 4.3.1.4.

1.16	Danny:	you attempted to do a search and it didnt come up with anything
1.17		(4.66) ((Hank types at the keyboard. Both he and Danny orient to the right monitor as it displays the result of the actions that Hank has initiated.))

Hank's actions take a turn in the transaction that Danny and Hank are producing. It is as if Hank says "I am doing it now, right here," yet in carrying out the work in front of his and Danny's gaze, the need for speaking is obviated—the work *signals itself*. This self-signaling of work is precisely what D2 exploits in telling D1 the *location* of her actions rather than describing the *content* of these actions in the Dourish and Belotti fragment.

This self-signaling, when perceivable features of a shared context are used semiotically, is what Roth (2004) calls a *perceptual gestalt*. "Perceptual gestalts in the environment available to co-participants in interaction … are sign forms that have no equivalent in the utterances or gestures but still constitute turn-taking units and therefore resources in an interaction" (p. 1039). Consistent with the study here, Roth shows through the fine-grained analysis of transactions among scientists in laboratories and field-settings that perceptual gestalts are a key feature of transactions that

need to be treated analytically on a par with speech and gesture to understand the order and rationality of situated human activity. When all of those who collaborate in the work share a set of perceptual gestalts, then work often proceeds with little speech as the participants focus on the action as it unfolds. But at points of trouble, when problems arise that need to be dealt with, the participants in a work setting will make the work visible and audible to one another.

Perceptual gestalts can be viewed as specific instances of what Goodwin (1994 p. 606) calls *professional vision*, the "socially situated, historically constituted body of practices through which the objects of knowledge which animate the discourse of a profession are constructed and shaped." Similarly, Kuhn (1962) discusses how scientists work within socially constituted *paradigms* that both represent and determine a community's ontology, and Fleck (1935, 1979) defines the "thought-styles" of a community as "the readiness for directed perception, with corresponding mental and objective assimilation of what has been so perceived." Further, Hanson describes that scientific observations are always theory-laden, so that seeing is never "objective" in a positivist sense, while Schütz (1962) describes the "typifications" in and through which the everyday world is socially constituted. Perceptual gestalts, while consistent with these prior conceptions, rather than being a broad characterization identifies specific instances of these socially constituted ways of seeing and categorizing the world.

There are two important conditions for perceptual gestalts to function semiotically, and when these are satisfied, the work can "go without saying" (Roth 2004 p. 1044). The first is that the interlocutors have to share sufficient expertise about the perceptual gestalt so as to distinguish the figure of the perceptual gestalt from the ground of all of their sensory input. Otherwise, the perceptual gestalt itself will be made visible for those with less expertise to learn from. "In a pedagogical situation, the perceptual gestalts actually become topics of the talk" (p. 1047). The second condition is socially recursive we-awareness: "co-participants may be oriented to the same section of the environment *and know that others are oriented in the same way* [emphasis added]" (p. 1041). Thus, for cooperating actors to exploit the semiotic possibilities of the work that is mutually heeded, they must not only orient to the same features of the environment, but they must each recursively know that the other person is so oriented, the central characteristic of we-awareness.

This leads to empirically testable predictions about the kinds of trouble that we can expect to be made visible and audible when one or the other of these conditions is not met. That is, we can expect that perceptual gestalts cannot be relied on for communicative purposes when either the interlocutors do not have sufficient shared competence in recognition and use of the gestalt, or the interlocutors are not mutually oriented and know that they are mutually oriented.

In the first case, this is exactly what we see in video data that we have in which a relative newcomer to BeamCoffer pairs with one of the old-timers—one session with Danny, and one with Hank. Although a detailed analysis of these sessions exceeds the scope of this paper, what we see is an order of magnitude more speech, fewer

segments of silence, and the requirement of several more minutes to become aligned at the start of each pairing session with the newcomer.

In the second case, we again turn to the data provided by Dourish and Belotti (1992). Consider turn 4.2 in the excerpt provided above, when D2 says "OK… Now I'll copy this; I'll cut this…" What is worth noting is that D2 *verbally announces what she is doing as she does it*. Were the copy and cut actions self-signaling as a perceptual gestalt, D2 would know it, and know that D1 knows it, and D1 would know that D2 knows it, etc. Hence, there would be no need for D2 to announce the very actions that she is performing. D2's words, then, carry out a semiotic function that is completely communicated by the work itself for Danny and Hank when they work side by side. It would, in fact, be strange if Hank were to announce, as D2 does, the work that he is doing as he does it with Danny sitting next to him, and Danny would likely wonder about the communicative purpose for Hank in stating the obvious. This fragment from Dourish and Belotti's designers makes audible precisely the semiotic content that Hank's mutually observed actions communicate.

5.3. We-awareness

We-awareness goes beyond aggregated I-awareness. Our claim is that the heeding strategy that *each person can at the same time see the computer monitor and what the other person is seeing* is not enough to account for Henry and David's accomplishment of cooperative work. What is further required is that David and Henry each can assume that the other is oriented in the same way, and they know that they can assume this, ad infinitum. That is, they engage in socially recursive inference and presuppose it of their partner in shaping and interpreting their actions and those that they perceive in their shared environment.

In embodied activity, acting is often viewed as separate from but complementary to perceiving. The acting body *displays* itself to others, the perceiving body *monitors* the display of others. "Displaying and monitoring are thus *complementary aspects* of the same coordinative practices [emphasis in original]" (Schmidt 2002b, p. 291). But in the pair programming work analyzed above, we see something occurring in addition when the two partners are both oriented to the same computer monitor and sitting side by side: monitoring is itself action, this action is perceivable by the other (i.e. it has display properties), this perception of the action of the other is perceivable, and so on.

5.3.1. *Basic and coordinating joint activity*

In discussing the communication that a transacting pair requires for cooperative work, Clark (2005) distinguishes between two kinds. "A joint activity can ordinarily be divided into two parts: the *basic joint activity* and the *coordinating joint activity* … People ordinarily cannot carry out a basic joint activity without communicating to coordinate it. At the same time, the reason they communicate about that activity is to coordinate on it" (p. 508). Using different words for the same distinction, Gutwin and

Greenberg (2002 p. 418) note that "[w]hen someone works alone in a workspace, their activities and their SA [situation awareness] involve only the workspace and the domain task. … In a collaborative situation, however, people must undertake another task, that of collaboration." Thus, for instance, when Danny says "You attempted to do a search and it didn't come up with anything," we can hear this as *coordinating joint activity* while Hank's operations on the computer in response can be seen as *basic joint activity.*

Mutually observed actions, whether of the person or computer, however, complicate and conflate these distinctions, in that under these conditions, all basic joint actions are at the same time coordinating actions, because of their signaling functions with respect to the actors. That is, the actions on and by the computer, when mutually (and recursively) observed by competent agents, serve to coordinate the activities of the pair in and of themselves without any requirement for the employment of additional speech or gesture. When these conditions are met, semiotic resources can be preserved without loss of coordination, saving time and likely reducing cognitive load. But when the conditions are not met, considerably more resources, especially speech, are likely to be required for coordinating work. "Although these nonverbal behaviors can be replaced with verbal substitutes, the substitutes take more time and effort (Brennan 1990)" (Kraut et al. 2002 p. 148).

To illustrate, consider the example of Hank and Danny, sitting side by side. Hank can see that Danny sees what Hank is doing. And similarly, Danny sees that Hank sees that Danny sees what Hank is doing, and so on, where this seeing allows recursive inference as many levels as the two require for their purposes. This socially recursive inference gives Danny and Hank sufficient common ground so that, for example, after 38.7 s of silently tracking the computer activity and one another's head orientations, Danny can say "better" and this can be mutually intelligible. In fact, "better" can be understood as the second part of an adjacency pair, where the first is constituted by the perceptual gestalts available to both. "Better," therefore, responds to what has been available to both participants, but "better" also is articulated *for* Hank, the recipient, and therefore is also owned by him. Unless the "better" come to be contested, it *is* the common ground. Danny knows that Hank has viewed what he has, and Danny also knows that Hank knows that Danny has viewed the same thing, so that "better" references what they have together grounded about the computation as it has executed. When you and I program together, I need to know not only where you are working and where you are looking, but also that you *know* that I know this, for as many recursive levels as we need so as to sufficiently ground in our current context. Because if you do not know that I know this, or I do not know that you do, then we cannot presuppose it as common ground that we share and exploit it in our communication. In other words, to do this type of work Danny and Hank need we-awareness.

5.3.2. *Tool support for we-awareness*

CSCW researchers have noted that the kind of tools described above, including screen sharing, real-time audio, and telepointers, have thus far failed to support

certain kinds of tightly coupled activity by actors cooperating at a distance. This is particularly the case with shared drawing tools, where gesture and embodiment figure prominently in collocated activity. Robertson points out that "[s]hared drawing applications have attempted to provide resources for both the perception of gesture by the use of telepointers and/or video links as well as the perception of all participants' drawing and writing actions in the shared drawing space" (1997 p. 148), echoed in (Kirk et al. 2005). These forms of tool support have also been used in distributed pair programming systems.

Stotts and Williams (Stotts and Williams 2002) describe adding video feeds to the above-described distributed pair programming systems that shows an image of the head of the other programmer projected in a small window on the left side of each programmer's computer monitor, obtained from a ceiling mounted video camera to the left of and pointing at the programmer's profile. This video-based distributed pair programming tool appears to support the aggregated heeding strategy. Each can see the computer display and the image of the other person's head orientation at the same time. The problem is that there is insufficient fidelity for the common ground condition: although both may see the other's head orientation and the computer display, they have insufficient knowledge that the other is actually doing this perceptual activity. But it is on just such recursive social inferences that Danny and Hank's "better" and "good" are grounded, and without these inferences, Danny and Hank will require additional communicative activity to achieve the grounding criterion. Stotts et al. (2004 p. 4) remark that in experiments with this system "[e]ach team turned off the video almost immediately."

As an analogy, imagine that we place a glass screen between Danny and Hank in the sessions above. Suppose as well that we tell each in private that the screen is a one-way mirror through which he can see but that is opaque to the other person. Because the glass is transparent, Danny can see Hank's head orientation and the computer monitor at the same time, and similarly for Hank. But Danny will not know that Hank sees the computer monitor and Danny's head orientation, since Danny believes that only he can see through the screen, and similarly for Hank. Thus, the aggregated heeding strategy is perfectly satisfied (each person can at the same time see the computer monitor and where the other person is oriented), but the common ground condition is not, since neither knows that the other person is similarly heeding. Thus, neither can rely upon this common ground in their activity. And this is the problem with all video systems for remote collaboration that provide images of each other's head orientation without the capability for knowing (inferring, presupposing) that the other person is actually perceiving this image.

One innovation that Stotts et al. (2004) subsequently developed shows promise for supporting this recursive inference on which we-awareness rests. Rather than projecting images of the other programmer into a separate window, they overlaid, onto the shared desktop showing the work, a transparent image of *both* programmers

stitched together from each computer's webcam as if the two programmers were sitting side by side at the workstation. This work, which they call *Facetop*, was inspired by the *ClearBoard* design for a shared, computer-mediated drawing medium by Ishii et al. (1992) based on a metaphor of "looking *through* and drawing *on* a big glass board" (p355). As the Facetop researchers state (p. 4), "The video image is …. tightly and seamlessly integrated with the shared workspace via transparency, thereby eliminating the 'dual' nature of video teleconferencing solutions. Users do not have to switch their attention from desktop, to video, back to desktop." Whether programmers using Facetop can actually perform the identified socially recursive inference is an empirical question. Although no such user tests have ever been carried out (Stotts, personal communication), this system nonetheless shows promise for supporting the socially recursive inference that collocated pair programmers are able to make.

5.3.3. *Short-circuiting embodied representations*

Technological solutions such as ClearBoard and Facetop provide a simulacrum of embodied co-presence by representing the visual aspects of the material body, thereby rendering it (and the actions that the body takes, including gaze and orientation) sensible to others. In the pair programming case, the programmers use one another's head orientation as a proxy for *where the eye is looking*, which is then used for inferring that each sees what the other sees. Common ground is achieved since both individuals see that the other sees their direction of gaze and what they are looking at simultaneously, in combination with an inference that the partner has similar psycho-perceptual abilities. Merleau-Ponty (2012) discusses the importance of this kind of reasoning about the psycho-perceptual equivalence of self and others, summarized by Robertson (2002 p. 306). "The fact that we are able to perceive at least some of our own bodily surfaces at the same time as we live our perceiving bodies … is the reason why we can recognize and understand others' actions by the same process that we shape our own actions for their interpretation by others." Meltzoff labels this the "like me" phenomenon, and it emerges at a very young age in people. "The child, even the newborn, processes the movements of other people and recognizes: 'that looks the way this feels' or 'those acts are like these acts.' The fact that others are seen as 'like me' provides an interpretive lens for infants' first social encounters" (Meltzoff 2011 p. 52). As we see here, "like me" reasoning continues to provide a fundamental basis for engaging in social transactions throughout the lifespan.

Identifying this socially recursive inference concerning where the self and the other are looking suggests a different design alternative than directly representing the gazing body; rather, the gaze *target* itself can be visually represented. This idea was first proposed by Velichkovsky (1995), who discusses the use of *visual cursors* using eye-tracking technology for use by two participants engaged in a collaborative problem-solving task. A visual cursor is a representation of where the other person is looking that is displayed, enabled by eye-tracking technology. For example, applying this technology to pair programming, the gaze target of each programmer could be directly represented such as in Figure 5, where a partially-transparent circle

```
public List<Role> getRole(User theUser)
{
        boolean isAuthor = false;
        boolean isSubprogramChair = false;

        final List<Role> roles = new ArrayList<Role>();
        if (theUser.equals(_programChair.getUser()))
                roles.add(_programChair);

        if (_registeredReviewers.contains(new Role(theUser, RoleType.REVIEWER)))
                roles.add(new Role(theUser, RoleType.REVIEWER));

        for (final Paper p : _registeredPapers)
        {
                if (theUser.equals(p.getAuthor().getUser()) &&
                                !isAuthor)
                {
                        roles.add(p.getAuthor());
                        isAuthor = true;
                }

                if ((p.getSubProgramChair() != null) &&
                        theUser.equals(p.getSubProgramChair().getUser()) &&
                        !isSubprogramChair)
```

Figure 5. Representing gaze, where each circle stands for the fixation point of one of the programmers

representing the visual fixation location (i.e. the gaze cursor) of each programmer is overlain on the displayed output. Each could then see the relationship of their own gaze to that of their partner and to what is "underneath" in the display, thereby perceiving the seeing and the seen in the same visual operation. As in the embodied, collocated case, this achieves common ground through the same kind of recursive "like me" inference process.

Neider et al. (2010) explicitly provide this "grounding" rationale in explaining the communicative efficiencies that cooperating, remote participants achieve on a location task within a visual scene using this dual eye-tracking/gaze-cursor technology. Gergle and Clark (2011) similarly demonstrate communicative efficiencies that collaborators using a similar technological system can achieve. "We also found that speakers used their addressee's gaze as an indication of attention. Speakers were more likely to use demonstrative deixis when they shifted the discourse focus to a new object and addressees were not looking at the speaker's intended referent. This highlights how speakers flexibly use different conversational resources to direct attention: when pairs have shared visual evidence, they rely less on language to communicate, but when they are not coordinated in their visual attention, specific referential forms can be used to direct attention" (pp442-3). Our account, then,

provides not only a grounding for these recent results in terms of the "we-awareness" that these technological developments facilitate, but also specifically highlights the recursive social inference that they support.

6. Conclusion

Awareness has been an important concept in CSCW since its formation as a distinct field of research and practice. When people collaborate, what is it that they need to be aware of in order to coordinate and carry out their interdependent activities? As Robertson (2002) points out, all attempts to support cooperative work through technological means embed a philosophical stance on what constitutes awareness and how it functions in joint human affairs. Considerable progress has been made in making such stances explicit, the way in which people immanently concern themselves in and with the world. Through their competence in cooperative work, they heed the publicly available resources in their environment, including those that they generate for themselves and others through their material embodiment.

As important as this conception is for structuring empirical research studies and as a basis for technological design, it is insufficient for characterizing the *shared* intentionality and socially recursive inference that philosophers, psychologists, and linguists are increasingly recognizing underwrite all forms of human cooperative activity. The design and interpretation of communication, as well as the interpretation of perceivable objects and events in the environment, depend on presuppositions that the actors in a setting have not only about their own intentions and the intentions of others, but about the we-intentions that they share. From the simplest gesture to the most complex utterance, interlocutors generate semiotic resources *for* others predicated on their assumptions about what the others know and their belief that these others are "like me" in their psychological and physical capabilities. And these same interlocutors interpret signs *from* others with the understanding of the assumptions underlying their generation.

Our goal in this paper has been to reconceptualize awareness as understood in CSCW by building from current conceptions to one of shared intentionality. We present an analysis of a case study of a setting in which a pair carries out tightly-coupled, collocated work mediated by computers, that of pair programmers working in situ in a software development organization. Through skilled practices of seeing, the programmers exploit the simultaneous perception of the computational work that is displayed on the computer monitor and their partner's orientation to the monitor to make the socially recursive inference that both are seeing the same thing at the same time, and that they know that they do so, ad infinitum. Not only do they use this inference to seamlessly coordinate and align their activity, they also recognize those times when they are not aligned and need to use explicit speech and gesture for building additional common ground about their shared intentions. Building on these case-based insights, we then suggest how attention to the socially recursive inference underlying the heeding of communicative actions and displays impacts the space of

design possibilities for computer-based technological support. Not only can this support simulate physical embodiment to enable participants to sense and be sensed simultaneously, it can also extend beyond the simulation of physical embodiment to a more direct representation of attention from which the recursive inferences can be made.

Acknowledgments

Josh Tenenberg thanks the University of Washington for granting a sabbatical leave during which much of this research was undertaken, as well as Professor Margaret-Anne Storey, the CHISEL group, and the Department of Computer Science at the University of Victoria for hosting him during this time. He also thanks Julia Schenk and Natalie Jolly for discussions during early phases of this research. David Socha thanks the Computing & Software Systems Division of the School of STEM at the University of Washington Bothell for funds to purchase recording equipment. He also was partially funded by a 2012–2013 Worthington Distinguished Scholar award from the University of Washington Bothell. Finally, all of the authors would like to thank the employees at BeamCoffer who generously offered to open their workspaces and practices to us, without whom none of this research would be possible.

References

"mutual, adj. and n." (2014). Retrieved November 11, 2014, from http://www.oed.com/

Baheti, Prashant, Edward F. Gehringer, and David Stotts. (2002). Exploring the Efficacy of Distributed Pair Programming. In *XP Universe* pp. 208–220

Beck, Kent. (1999). *Extreme Programming Explained: Embrace Change*. Reading, MA: Addison-Wesley Professional.

Bratman, Michael. (1992). Shared Cooperative Activity. *The Philosophical Review*, vol. 101 no. 2, pp. 327–341.

Button, Graham, and Wes Sharrock. (2000). Design by problem-solving. In P. Luff, J. Hindmarsh, and C. Heath (Eds.): *Workplace Studies Recovering Work Practice and Informing System Design* pp. 46–67. Cambridge, UK: Cambridge University Press

Cadoz, Claude. (1994). *Les réalités virtuelles*. Paris, France: Flammarion.

Calefato, Fabio, and Filippo Lanubile. (2012). Augmenting Social Awareness in a Collaborative Development Environment. *2012 5th International Workshop on Co-Operative and Human Aspects of Software Engineering (CHASE)*, pp. 13–15

Clark, Herbert H. (1993). *Arenas of language use*. Chicago, IL USA: University of Chicago Press.

Clark, Herbert H. (2003). Pointing and Placing. In S. Kita (Ed.): *Pointing: Where Language, Culture, and Cognition Meet* pp. 243–68. Mahwah, NJ USA: Lawrence Erlbaum Associates

Clark, Herbert H. (2005). Coordinating with each other in a material world. *Discourse Studies*, vol. 7 no. 4–5, pp. 507–525.

Clark, Herbert H., and Susan E. Brennan. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, and S. D. Teasley (Eds.): *Perspectives on socially shared cognition* Vol. 13, pp. 127–149. American Psychological Association

Cohn, Marisa Leavitt, Susan Elliott Sim, and Charlotte P. Lee. (2009). What Counts as Software Process? Negotiating the Boundary of Software Work Through Artifacts and Conversation. *Computer Supported Cooperative Work*, vol. 18, pp. 401–443.

Convertino, Gregorio, Helena M. Mentis, Aleksandra Slavkovic, Mary Beth Rosson, and John M. Carroll. (2011). Supporting Common Ground and Awareness in Emergency Management Planning: A Design Research Project. *ACM Transactions on Computer-Human Interaction*, vol. 18 no. 4

Dajda, Jacek, and Grzegorz Dobrowolski. (2007). How to build support for distributed pair programming. In *Proceedings of the 8th international conference on Agile processes in software engineering and extreme programming (XP'07)* pp. 70–73. Springer-Verlag

Dourish, Paul, and Victoria Belotti. (1992). Awareness and Coordination in Shared Workspaces. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work* pp. 107–114

Dreyfus, Hubert. (1991). *Being-in-the-World: A Commentary on Heidegger's Being and Time, Division I*. Cambridge, MA USA: MIT Press.

Fleck, Ludwik. (1935). *Entstehung und Entwicklung einer wissenschaftlichen Tatsache. Einfuhrung in die Lehre vom Denkstil und Denkkollectiv.* (S. und Co., Ed.). Basel, Switzerland

Fleck, Ludwik. (1979). *Genesis and development of a scientific fact*. (T. J. Trenn & R. K. Merton, Eds.). Chicago, IL USA: University of Chicago Press

Flor, Nick V. (2006). Globally distributed software development and pair programming. *Communications of the ACM*, vol. 49 no. 10, pp. 57

Gergle, Darren, and Alan T. Clark. (2011). See what i'm saying?: using Dyadic Mobile Eye tracking to study collaborative reference. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work (CSCW'11)* pp. 435–444.

Gilbert, Margaret. (1989). *On Social Facts*. London, UK: Routledge & Kegan Paul Ltd.

Gilson, Richard D., Daniel J. Garland, and Jefferson M. Koonce (Eds.). (1994). *Situational Awareness in Complex Systems*. Daytona Beach, FL USA: Embry-Riddle Aeronautical University Press

Goodwin, Charles. (1994). Professional Vision. *American Anthropologist*, vol. 96 no. 3, pp. 606–633.

Goodwin, Charles. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics*, vol. 32, pp. 1489–1522.

Goodwin, Charles, and Marjorie Harness Goodwin. (1996). Seeing as a Situated Activity: Formulating Planes. In Y. Engeström and D. Middleton (Eds.): *Cognition and Communication at Work.* Cambridge, UK: Cambridge University Press.

Goodwin, Charles, and John Heritage. (1990). Conversation Analysis. *Annual Review of Anthropology*, vol. 19, pp. 283–307.

Grice, Paul. (1975). Logic and conversation. In P. Cole and J. Morgan (Eds.): *Syntax and Semantics, 3: Speech Acts*. New York: Academic Press.

Gross, Tom. (2013). Supporting Effortless Coordination: 25 Years of Awareness Research. *Computer Supported Cooperative Work*, vol. 22, pp. 425–474.

Gutwin, Carl, and Saul Greenberg. (2002). A descriptive framework of workspace awareness for real-time groupware. *Computer Supported Collaborative Work*, vol. 11 no. 3/4, pp. 411–446.

Gutwin, Carl, and Saul Greenberg. (2001). The Importance of Awareness for Team Cognition in Distributed Collaboration. Report 2001-696-19, Dept Computer Science, University of Calgary, Alberta, CANADA.

Hanks, Brian. (2008). Empirical evaluation of distributed pair programming. *International Journal of Human-Computer Studies*, vol. 66, pp. 530–544.

Heath, Christian, and Paul Luff. (2000). *Technology in Action*. Cambridge U.K.; New York NY USA: Cambridge University Press

Heath, Christian, Marcus Sanchez Svensson, Jon Hindmarsh, Paul Luff, and Dirk Vom Lehn. (2002). Configuring awareness. *Computer Supported Cooperative Work*, vol. 11 no. 3/4, pp. 317–347.

Heidegger, Martin. (1962). *Being and Time*. (J. Macquarrie & E. Robinson, Trans.). New York, NY USA: Harper and Row

Holzkamp, K. (1983). *Grundlegung der Psychologie*. Frankfurt/M, Germany: Campus

Hutchins, Edwin. (1993). Learning to Navigate. In S. Chaiklin and J. Lave (Eds.): *Understanding Practice: Perspectives on Activity and Context* pp. 35–63. Cambridge University Press.

Ishii, H., M. Kobayashi, and Jonathan Grudin. (1992). Integration of inter-personal space and shared workspace: ClearBoard design and experiments. In *Proceedings of ACM Conference on Computer Supported Cooperative Work* pp. 33–42

Jones, Capers. (1991). *Applied software measurement: assuring productivity and quality.* New York: McGraw-Hill.

Jordan, Brigette Brigitte, and Austin Henderson. (1995). Interaction Analysis: Foundations and Practice. *The Journal of the Learning Sciences*, vol. 4 no. 1, pp. 39–103.

Kirk, David, Andy Crabtree, and Tom Rodden. (2005). Ways of the Hands. In *Proceedings of the Ninth European Conference on Computer-Supported Cooperative Work* pp. 1–22

Kirk, David, Tom Rodden, and Stanton Danaë Fraser. (2007). Turn It This Way: Grounding Collaborative Action with Remote Gestures. In *Proceedings of the 25th annual ACM conference on Human factors in computing systems - CHI'07* pp. 1039–1048.

Kopelman, J. Mark Weber, and David M. Messick. (2002). Factors Influencing Cooperation in Commons Dilemmas: A Review of Experimental Psychological Research. In E. Ostrom, T. Dietz, N. Dolšak, P. C. Stern, S. Stonich, and E. U. Weber (Eds.): *The Drama of the Commons*. Washington, DC USA: National Academies Press

Koschmann, Timothy, and Curtis D. LeBaron. (2003). Reconsidering Common Ground: Examining Clark's Contribution Theory in the OR. In *Proceedings of the Eighth European Conference on Computer Supported Cooperative Work, ECSCW'03*.

Kraut, Robert, Susan Fussell, Susan E. Brennan, and Jane Seigel. (2002). Understanding Effects of Proximity on Collaboration: Implications for Technologies to Support Remote Collaborative Work. In P. Hinds and S. Kiesler (Eds.): *Distributed Work* pp. 137–162. Cambridge, MA USA: MIT Press

Kuhn, Thomas S. (1962). *The structure of scientific revolutions*. Chicago, IL USA: University of Chicago Press.

Larkin, Michael, Virginia Eatough, and Mike Osborn. (2011). Interpretative phenomenological analysis and embodied, active, situated cognition. *Theory & Psychology*, vol. 2 no. 13, pp. 318–337.

Leont'ev, A. N. (1978). *Activity, Consciousness, and Personality.* (M. J. Hall, Trans.). Upper Saddle River, NJ USA: Prentice Hall

Luff, Paul, Christian Heath, and Marcus Sanchez Svensson. (2008). Discriminating Conduct: Deploying Systems to Support Awareness in Organizations. *International Journal of Human-Computer Interaction*, vol. 24 no. 4, pp. 410–436.

Meltzoff, Andrew N. (2011). Social cognition and the origins of imitation, empathy, and theory of mind. In U. Goswami (Ed.): *The Wiley-Blackwell handbook of childhood cognitive development* pp. 49–75. Malden, MA USA: Wiley-Blackwell

Merleau-Ponty, Maurice. (1962). *Phenomenology of Perception*. (C. Smith, Trans.). London, UK: Routledge & Kegan Paul Ltd.

Merleau-Ponty, Maurice. (2012). *Phenomenology of Perception*. (D. A. Landes, Trans.). New York, NY USA; Abingdon, UK: Routledge

Meshcheryakov, A. (1979). *Awakening to Life: On the Education of Deaf-blind Children in the Soviet Union*. Moscow, USSR: Progress Publishers.

Neider, Mark B., Xin Chen, Christopher A. Dickinson, Susan E. Brennan, and Gregory J. Zelinsky. (2010). Coordinating spatial referencing using shared gaze. *Psychonomic Bulletin & Review*, vol. 17 no. 5, pp. 718–724.

Ostrom, Elinor. (2005). *Understanding institutional diversity.* Princeton, NJ USA: Princeton University Press.

Plonka, Laura. (2012). *Unpacking Collaboration in Pair Programming in Industrial Settings*. Doctoral dissertation, The Open University.

Pomerantz, Anita. (1984). Agreeing and disagreeing with assessments: some features of preferred/dispreferred turn shapes. In J. M. Atkinson and J. Heritage (Eds.): *Structures of Social Action: Studies in Conversation Analysis*. Cambridge, UK: Cambridge University Press.

Robertson, Toni. (1997). *Designing Over Distance: A Study of Cooperative Work, Embodied Cognition and Technology to Enable remote Collaboration*. Doctoral dissertation, University of Technology, Sydney.

Robertson, Toni. (2002). The public availability of actions and artifacts. *Computer Supported Cooperative Work*, vol. 11 no. 3/4, pp. 299–316.

Rooksby, John, Mark Rouncefield, and Ian Sommerville. (2009). Testing in the Wild: The Social and Organisational Dimensions of Real World Practice. *Computer Supported Cooperative Work*, vol. 18, pp. 559–580.

Roth, Wolff-Michael. (2003). From epistemic (ergotic) actions to scientific discourse: The bridging function of gestures. *Pragmatics & Cognition*, vol. 11 no. 1, pp. 141 – 170.

Roth, Wolff-Michael. (2004). Perceptual gestalts in workplace communication. *Journal of Pragmatics*, vol. 36 no. 6, pp. 1037–1069.

Roth, Wolff-Michael. (2011). *Geometry as objective science in elementary classrooms: Mathematics in the flesh*. New York, NY USA: Routledge.

Roth, Wolff-Michael. (2013). *What More in/for Science Education: An Ethnomethodological Perspective*. Rotterdam/Boston/Taipei: Sense Publishers.

Roth, Wolff-Michael. (2014a). Science language Wanted Alive: Through the dialectical/dialogical lens of Vygotsky and the Bakhtin circle. *Journal of Research in Science Teaching, 51*, 1049–1083.

Roth, Wolff-Michael. (2014b). Working out the interstitial and syncopic nature of the human psyche: On the analysis of verbal data. *Integrative Psychological and Behavioral Science*, vol. 48, pp. 283–298.

Roth, Wolff-Michael, and Alfredo Jornet. (2013). Situated cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, vol. 4 no. 5, pp. 463 – 478.

Roth, Wolff-Michael, and Kenneth Tobin. (2010). Solidarity and conflict: Aligned and misaligned prosody as a transactional resource in intra- and intercultural communication involving power differences. *Cultural Studies of Science Education*, vol. 5, pp. 805–847.

Ryle, Gilbert. (1949). *The Concept of Mind*. London, UK: Hutchinson.

Sacks, Harvey. (1992). *Lectures on conversation 1964–1972, Volume I & II*. (G. Jefferson, Ed.). Malden, MA USA: Blackwell Publishing

Sacks, Harvey, Emanuel Schegloff, and Gail Jefferson. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, vol. 50 no. 4, pp. 696–735.

Salinger, Stephan, Christopher Oezbek, Karl Beecher, and Julia Schenk. (2010). Saros: an eclipse plug-in for distributed pair programming. In *Proceedings of the 2010 ICSE Workshop on Cooperative and Human Aspects of Software Engineering* pp. 48–55

Schmidt, Kjeld. (1997). Of maps and scripts: The status of formal constructs in cooperative work. In *GROUP'97, ACM Conference on Supporting Group Work* pp. 138–147.

Schmidt, Kjeld. (2002a). Remarks on the complexity of cooperative work. *Revue Des Sciences et Technologies de L'information*, vol. 16 no. 4–5, pp. 443–483.

Schmidt, Kjeld. (2002b). The Problem with Awareness. *Computer Supported Cooperative Work*, vol. 11, pp. 285–298.

Schmidt, Kjeld. (2011). *Cooperative Work and Coordinative Practices - Contributions to the Conceptual Foundations of Computer-Supported Cooperative Work (CSCW)*. Heidelberg, Germany: Springer-Verlag.

Schummer, T., and Stephan Lukosch. (2009). Understanding tools and practices for distributed pair programming. *Journal of Universal Computer Science*, vol. 15 no. 16, pp. 3101–3125.

Schütz, Alfred. (1962). Common-sense and scientific interpretations of human action. In *Alfred Schütz Collected Papers, Volume 1: The Problem of Social Reality* pp. 3–47. The Hague: Martinus Nijhoff.

Searle, John. (1990). Collective Intentions and Actions. In P. R. Cohen, J. Morgan, and M. Pollack (Eds.): *Intentions in Communication*. Cambridge, MA USA: MIT Press.

Searle, John. (1995). *The Construction of Social Reality*. New York, NY USA: The Free Press.

Sharp, Helen, and Gina Venolia. (2009). Cooperative and Human Aspects of Software Engineering. *IEEE Software*, vol. 26 no. December, pp. 2009–2011.

Sharrock, Wes, and Graham Button. (2011). Engineering Investigations: What is Made Visible in Making Work Visible. In M. Szymanski and J. Whalen (Eds.): *Making Work Visible: Ethnographically Grounded Case Studies of Work Practice* pp. 34–50. Cambridge, UK: Cambridge University Press

Spector, Robert H. (1990). Chapter 116: Visual Fields. In H. K. Walker, W. D. Hall, and J. W. Hurst (Eds.): *Clinical Methods: The History, Physical, and Laboratory Examinations.* 3rd ed. Boston, MA USA: Butterworths.

Stotts, David, Jason McC. Smith, and Karl Gyllstrom. (2004). Support for distributed pair programming in the transparent video facetop. In *Proceedings of the Fourth Conference on Extreme Programming and Agile Methods— XP/Agile Universe* pp. 92–104.

Stotts, David, and Laurie Williams. (2002). *A Video-enhanced Environment for Distributed Extreme Programming*. Raleigh, North Carolina USA: Internal Report, North Caroline State University.

Suchman, Lucy. (1987). *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge, MA USA: Cambridge University Press.

Tomasello, Michael. (2009). *Why we cooperate*. Cambridge, MA USA, MA: MIT Press.

Tomasello, Michael. (2014). *A Natural History of Human Thinking*. Cambridge, MA USA.; London, England: Harvard University Press

Tomasello, Michael, and Hannes Rakoczy. (2003). What Makes Human Cognition Unique? From Individual to Shared to Collective Intentionality. *Mind & Language*, vol. 18 no. 2, pp. 121–147.

Tomasello, Michael, M. Carpenter, J. Call, T. Behne, and H. Moll. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, vol. 28, pp. 675–691.

Tomasello, Michael, Alicia P. Melis, Claudio Tennie, Emily Wyman, and Esther Herrmann. (2012). Two Key Steps in the Evolution of Human Cooperation: The Interdependence Hypothesis. *Current Anthropology*, vol. 53 no. 6, pp. 673–692.

Velichkovsky, Boris M. (1995). Communicating attention: Gaze position transfer in cooperative problem solving. *Pragmatics & Cognition*, vol. 3 no. 2, pp. 199–223(25).

Vygotskij, L. S. (2005). *Psixologija razvitija čeloveka [Psychology of human development]*. Moscow, Russia: Eksmo.