Title
Coordinating turning while walking and talking

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 35(35)

ISSN
1069-7977

Authors
Mayor, Eric
Bangerter, Adrian

Publication Date
2013

Peer reviewed

Coordinating turning while walking and talking

Eric Mayor (eric.mayor@unine.ch)
Institute of Work and Organizational Psychology, University of Neuchâtel
Emile-Argand 11, 2000 Neuchâtel, Switzerland

Adrian Bangerter (adrian.bangerter@unine.ch)
Institute of Work and Organizational Psychology, University of Neuchâtel
Emile-Argand 11, 2000 Neuchâtel, Switzerland

Abstract

Few studies have investigated multitasking in joint actions, especially two joint actions performed by two people together and coordinated via multimodal communication. We investigate the case of two people walking and talking together, a common combination of joint actions. In an experiment, pairs talked together in four varying conditions of mobility. A narrator told a story to a partner. They did this while either standing immobile, walking along a straight-line itinerary, or walking along a complex itinerary featuring several turns. They also completed a walking task along a complex itinerary without having to tell a story. One person (the navigator) was also entrusted with a map of the itinerary. We analyzed how participants coordinated turning while telling a story. Narrators relied more on verbal means to signal turning, and were more distracted during the turn, leading to more repetition of story-related content.

Keywords: conversation, coordination, walking, multimodal communication, joint action, collaboration, multitasking.

Multimodal Coordination of Concurrent Joint Actions

Multitasking, or the concurrent performance of two different tasks, is common in everyday life. An important question concerns the effect of multitasking on task performance. Research on multitasking has revealed much about the basic cognitive processes involved, showing that sharing processing resources (attention, working memory, and executive control) between multiple tasks can impair performance (Salvucci & Taatgen, 2011). Much of this research, however, has focused on multitasking behavior of individuals engaged in solitary tasks. Some research focuses on situations where people coordinate concurrent joint actions (e.g., either an individual action and a joint action or two joint actions). For example, Fussell, Kiesler, Setlock, Scupelli, and Weisband (2004) investigated how people coordinated two projects, each one with a different partner, face-to-face and via instant messaging. But few studies investigate the role of dialogue in coordinating concurrent joint actions. This is a significant oversight, because dialogue (which involves both verbal and nonverbal acts, i.e., *multimodal* communication; Stivers & Sidnell, 2005) is the commonest means of coordinating joint action (Clark, 1996).

Investigations into the role of multimodal dialogue in coordinating multiple joint actions can significantly expand cognitive science research on multitasking. Recognizing how processing resources are distributed among multiple individuals and coordinated over multiple communicative modalities challenges existing cognitive theories on multitasking. As we will see, investigating such phenomena requires theories about coordinating meaning and identities in interaction.

What coordination problems arise when people perform concurrent joint actions, and how do people use multimodal communication to solve these problems? An initial investigation of this issue was proposed by Chevalley and Bangerter (2010). They used Clark's (1996) theory of language use to propose a model of how people suspend a joint action they are doing together in response to an interruption, and how they reinstate those actions after the interruption. Participants have to coordinate on at least three aspects in switching from one joint action to another. First, when reinstating a joint action after a suspension, participants have to update their common ground (Clark, 1996) about the state of the action. They do this by talking together about where they were in the action. Second, they have to attend to their partners' *face needs*. According to politeness theory (Brown & Levinson, 1987), people have a right to positive consideration by others (positive face) as well as a right to act freely without being unduly imposed on (negative face). Suspending a joint action and making one's partner wait while one does something else constitutes a threat to negative face. To mitigate this threat, participants engage in politeness like warning about an interruption, asking permission to suspend, minimizing its duration (*just a sec*) or apologizing. Third, coordinating responses to an interruption raises the question of a division of labor among interaction partners. For example, only one participant in a joint action may be the target of an interruption, leaving the other participant free to keep the current state of the action in memory. The non-interrupted participant may then play a crucial role in reconstructing the state of the action once the interruption is over. Indeed, asymmetries in conversational roles or access to privileged knowledge affect the way partners coordinate suspending and reinstating joint actions (Bangerter, Chevalley, & Derouwaux, 2010).

Here we pursue this line of inquiry but focus on the case where two people accomplish two joint actions concurrently with each other (rather than suspending one joint action for a longer period of time in order to engage in another one possibly involving another person). In such a case, conflicts between resources used for one task but required for another may arise. Multimodal communication is a potential way of circumventing this “bottleneck”, because communicating about one joint action in a different modality (e.g., via gestures) might leave the primary modality (e.g., talk) undisturbed. Another way of circumventing the bottleneck is to distribute task components among different individuals (Hutchins, 1995). In doing so, participants in multiple concurrent joint actions minimize overall collaborative effort (Clark, 1996). More generally, in coordinating multiple concurrent joint actions, participants respond to two fundamental imperatives of conversation (Enfield, 2006). An *informational* imperative requires participants to coordinate joint understanding of both actions (e.g., where they are in a narrative, when they are going to turn a corner), and an *affiliational* imperative requires them to manage each other’s identities and commitments to the joint action (e.g., not interrupting a speaker at an interesting point in a story). We apply theories of conversation as joint action to explain processes occurring in multimodal coordination of concurrent joint actions.

Walking While Talking

We report initial findings from an experimental study investigating how two people coordinate two concurrent joint actions, namely talking together while walking together. We chose walking and talking because it is a commonly occurring combination in everyday life. Many everyday conversations take place in situations of mobility. For example, hospital personnel spend substantial amounts of time engaging in various activities while walking (Bardram & Bossen, 2005).

Talking together is a common joint action that is coordinated through a variety of channels, including speech, paraverbal information, gaze, gesture, body posture and so on. Depending on the type of conversation, participants may occupy different roles that constrain their relevant contributions. Of course, talking together has been largely studied in various disciplines (Sacks, Schegloff, & Jefferson, 1974, Clark, 1996), but comparatively little is known about how conversation is coordinated with other, non-linguistic joint actions.

Walking together is also a common joint action where partners must coordinate walking speed and posture in order to position themselves abreast of each other. Synchronizing gait requires coordination via tactile (hand-holding) or visual signals (Zivotofsky & Hausdorff, 2007). In some cases, when walking constitutes a means of locomotion to a particular place known to only one of the partners, roles may also emerge (i.e., one person using a map). Indeed, even transitory forms of collective mobility like crossing a street as a group when the traffic light for pedestrians is green require coordination (Relieu, 2008).

Walking normally requires few cognitive resources, and people are typically able to walk and do something else at the same time. But there are measurable decrements in task performance in such cases. For example, older adults are less able to memorize while walking (Lindenberger, Marsiske & Baltes, 2000). Also, adults who answer questions while walking are less fluent than while stationary (Kemper, Herman, & Lian, 2003) Another study (Yatani & Truong, 2009) found that users of handheld devices are more effective when standing than when walking. These studies fall short of studying true joint actions because they do not investigate interactive conversation. However, they are relevant for understanding walking performed in conjunction with other actions, and suggest that the small decrements in performance could be easily increased by making walking more difficult (e.g., by having participants navigate a complex itinerary using a map rather than just walking a predetermined path). Thus, walking constitutes a convenient and malleable candidate task to investigate in conjunction with talking.

Our Experiment

In our experiment, pairs of participants were videotaped while talking together in four within-subjects conditions of varying mobility (the *Task* variable) designed to instantiate different combinations of concurrent demands related to walking and talking (Table 1). The talking task involved one person (the *narrator*) telling a story to the other (the *partner*). Participants kept these roles for the duration of the study. In the *talk-only* condition, pairs were standing immobile while the narrator told the story. In the *talk-and-walk* condition, they walked together along a straight-line itinerary which was indicated on a map while the narrator told the story. In the *talk-and-navigate* condition, they walked together along a complex itinerary (i.e., featuring five turns) which was indicated on a map while the narrator told the story. In the *navigate-only* condition, they walked together along a similarly complex itinerary (i.e., also featuring five turns) which was indicated on a map but could talk about whatever they wanted, thus creating a situation where navigation is clearly prioritized.

Table 1. Demands of talking and walking instantiated in four within-subjects conditions.

	Talking Demands	Walking Demands
Talk Only	High	None
Talk and Walk	High	Low
Talk and Navigate	High	High
Navigate Only	Low	High

In addition, either the narrator or the partner was entrusted with the responsibility of making sure the pair followed the itinerary correctly. The person responsible (hereafter the *navigator*) was given the map. This constituted a between-subjects variable.

Thus, the design of the experiment was a 4 (Task, within-subjects) X 2 (Navigator, between-subjects) design. In such a setting, it is possible to investigate many interesting questions. For example, the coordination of story-telling involves the narrator regularly seeking a back-channel response from the partner. This is often done via gaze (Bavelas, Coates, & Johnson, 2002). If the partner is distracted and thus kept from producing back-channel responses, the quality of the story suffers (Bavelas, Coates, & Johnson, 2000). However, when walking and talking, gaze may not be as freely available for this purpose as when people are talking without moving. The effect of walking on gaze allocation and therefore on story-telling coordination via back-channels can be investigated by comparing the talk-only condition with the other conditions. Other comparisons are possible, for example comparing the talk-and-navigate condition with the navigate-only condition allows investigating to what extent talking may interfere with a navigational task, with navigational performance being measured by changes in walking speed (e.g., slowing down or stopping) or by errors (e.g., wrong turns).

In this paper, our analysis focuses on how participants coordinate turning to the left or to the right according to the itinerary while talking. Turning while talking is a good example of how an acute coordination demand may emerge from one joint action, thereby jeopardizing coordination of the other joint action. In our experiment, the responsibility for navigating was often implicitly entrusted to the navigator, who was the only participant who had easy visual access to the map. Thus, turning was typically coordinated via some kind of signal from navigators to the other participant. There are several ways to do this. Navigators might tell other participants to turn, for example by uttering *we're going to turn to the right*. Or they might point in the direction of the turn. They might also swivel their gaze in the direction of the turn, or nudge or push their partner, or use a combination of several signals. Some pairs even managed to turn without any visible or audible coordination signals (albeit quite rarely). How might participants decide to coordinate a turn? When narrators are navigators, they have the floor, because they are responsible for telling the story. Thus, it seems easier for them to signal the turn via verbal means. On the other hand, when partners are navigators, they must interrupt the narrator and gain the floor if they want to signal the turn verbally. This is a potential threat to the narrator's face (Bangerter, Chevalley, & Derouwaux, 2010). If, as predicted by joint action theories of conversation, participants deal with this problem by distributing collaborative effort across modalities and by a distribution of labor, we would expect partners as navigators to rely relatively less on verbal means to signal turning than narrators as navigators.

To test this possibility, we investigated the effect of the Task and Navigator variables on the coordination of turning. For each of the five turns in the talk-and-navigate condition for each pair, we coded what kind of verbal or nonverbal means they used to coordinate the turn. We compared this data with the verbal and nonverbal means used to coordinate turning in the navigate-only condition. Because there are no narrator and partner roles in the navigate-only condition, it serves as a baseline for comparison with the effect of roles in the talk-and-navigate condition.

We also investigated the effect of the Task and Navigator variables on the coordination of storytelling. When narrators are navigators, they may be more distracted when they have to both communicate about the turn and keep track of the story they are telling. This might make participants more likely to lose track of the story, and thus more likely that some utterance relative to the story will have to be repeated after the turn as a means of reconstructing the story line (Chevalley & Bangerter, 2010).

Method

Participants

Eighty people (46 women and 34 men) participated in 40 pairs. Pairs were composed irrespective of gender. Participants were native French speakers and did not know each other before the study.

Procedure

We video-recorded each pair in one static and three mobile conditions. In all conditions, participants were also equipped with audio recorders and tie-clip microphones. In the talk-only condition, participants were filmed with a hand-held video camera from a distance of several meters. In the three mobile conditions, participants walked abreast. They were filmed frontally with a device consisting of either a GoPro Hero2 camera or a Contour HD camera attached to a perch that was held by the experimenter who walked about 1.5 m behind the pair. The perch extended over the heads of the pair (see Figure1). It was just above and the out of their field of vision when they looked ahead. The experimenter calibrated his walking speed to the participants' in order to maintain the camera at a constant distance from them. The perch also featured a supplementary backup audio recorder attached above the participants' heads. In this way, the setup allowed frontal mobile videotaping of the participants from above their heads to below their knees (Figure 2).

Figure 1: Setup of portable videocamera perch.

Twenty ordered combinations of the four conditions were randomly computed and randomly assigned to pairs in each between-subjects condition (the same combinations were used in both conditions). Pairs performed the tasks in the order thereby defined.

In the walking conditions, participants followed an itinerary using a map, responsibility for the navigation being randomly assigned to the narrator or partner before the experiment. Participants were asked to navigate from a starting point to a precisely marked end point. Thus, even straight-line itineraries required some monitoring on the part of the navigator to avoid undershooting or overshooting the end point. All itineraries had a total length of approximately 400 meters. Recordings took place outdoors in a quiet urban area.

Figure 2: Still pictures of two pairs (in both cases, the narrator is the navigator and is on the left). Bottom picture: The narrator is initiating a turn by gesturing.

Data preparation

Video was synchronized with the sound of the two audio recorders (on a separate track) and a file was produced per condition for each group. A video clip of each turn was prepared. Clips started approximately 15 seconds before the initiation of the turn and lasted 30 seconds.

Based on a viewing of each clip, a detailed qualitative description of how each pair coordinated each turn was written by the first author. The description featured a sequential list of the circumstances of the turn, as well as any visible or audible behavior dedicated to coordinating the turn, including specifications of which participant was on the inside of the turn, descriptions of gestures (e.g., pointing), verbatim transcription of any utterances or the direction of gaze.

A research assistant then coded each description on the following variables:

- *Who* produced a signal (narrator or partner).
- *When* it was produced (before, during or after the turn)
- *The signal produced* (look at map, look at other participant, look in the direction of the turn, look elsewhere, point in the direction of the turn, point on the map, other gesture, give directions verbally, request help, agree)
- *Repetitions* of previous story content

Interrater agreement was assessed by having two coders independently code 25 turns. Cohen's kappa indicated excellent agreement (all kappas > .90).

The individual turn-coordination signals were grouped together to compute frequencies with which three types of signals were produced: gaze, gesture and utterance. The number of repetitions per turn was also computed.

Results

Pairs took the same amount of time to complete the task in all four conditions, Wilks' lambda = .930, $F(3,37) = .922$, $p = .44$, ($M = 297.5$ s, $SD = 65.7$ s).

Because Task is a within-subjects variable, we performed repeated-measures analyses with the frequencies (by turn) of gaze, gesture, utterance and repetition as dependent variables. Because turns are nested within groups and the dependent variables are count data, we ran mixed model Poisson regressions in R 3.0. These analyses take into account the random effects of pairs. The independent variables were Navigator role and Task, which were entered in that order in the models, prior to the interaction term. Independent variables were dummy coded (0 vs 1). Categories coded 0 were Partner for the Navigator variable and Talk-and-navigate for the Task variable. The models were fitted by the Laplace approximation. Table 2 shows the means for each dependent variable as a function of Task and Navigator role. In what follows, *b* coefficients for each effect represent natural-log-transformed values.

Table 2: Mean frequencies (standard deviations) of gaze, gestures, utterances, and repetitions by Task and Navigator role per turn.

	Talk-and-navigate		Navigate-only	
	Narrator	Partner	Narrator	Partner
Gaze	4.50 (2.52)	4.36 (2.74)	3.92 (2.57)	4.13 (2.29)
Gesture	1.34 (2.36)	0.87 (1.03)	0.81 (1.30)	0.69 (1.12)
Utterance	1.63 (2.92)	0.58 (1.24)	1.56 (2.34)	1.36 (2.26)
Repetition	0.15 (0.38)	0.03 (0.17)	0.10 (0.30)	0.12 (0.41)

Gaze is used frequently in coordinating turning. While gaze shifts might be primarily produced by participants to steer their own individual walking trajectory, they might also attract the attention of the other participant and thus serve as an unintended cue that a turn is imminent. Pairs gazed marginally less in the navigate-only condition than in the talk-and-navigate condition (b = -.12, SE = 0.07, p = 0.07). Navigator role was not a significant predictor of gaze. It is worth noting that this model does not fit the data significantly better than a null model (*deviance* = 4, df = 3, *ns*). (Differences from the null models are significant for all other dependent variables.)

Gestures were used regularly, albeit less often than gaze. Pairs gestured marginally less when the partner was responsible for the itinerary than when the narrator was (b = -0.36, SE = 0.20, p = 0.07). In the navigate-only condition, pairs gestured less than in the talk-and-navigate condition (b = -0.48, SE = 0.14, p = 0.0007). As expected, pairs in the navigate-only condition used less utterances to coordinate turning than did pairs in the talk-and-navigate condition (b = -1.06, SE = 0.29, p = 0.0003). The interaction of task and navigator was also significant (b = .86, SE = *0.19*, p < 0.0001)*:* In the talk-and-navigate condition, pairs discussed the navigation task more when the narrator was responsible for navigation than when the partner was. On the contrary, in the navigate-only condition, pairs discussed the navigation task equally often, irrespective of navigator role.

Figure 4: Example of a progressive breakdown in the story following a missed turn.

Utterances related to turning included directions but also expressions of uncertainty, like *I just need to look* or *I think that's it*, as well as occasional requests for assistance, which sometimes could completely override the narrative activity. In one exceptional case (depicted in Figure 4), the narrator progressively realizes she is lost, first interrupting her story by saying *I don't understand where to go anymore* while pointing vaguely in the direction of the turn. She then looks at her partner and laughs, and then asks her *can you help me*, while showing the map to her partner. All the while, the pair is walking straight ahead without slowing down. Subsequent to the frames shown in Figure 4, the pair will slow down and come to a complete stop while the partner explains to the narrator where to go. Only once they have corrected their trajectory will the narrator resume her story. This example illustrates the complex interplay of the multimodal signals produced (verbal utterances, gaze, pointing, and showing the map). It also illustrates a momentary but complete breakdown in one task (talking) when coordination requirements of the other task (walking) briefly overwhelm participants' available resources.

Repetitions of story-related content were infrequent. When they did occur, it was mostly the last utterance before the turn that was repeated immediately after the turn was complete. Nonetheless, repetitions of story-related utterances were less frequent when the partner was responsible for navigation than when the narrator was (b = -1.60, SE = 0.71, p = 0.02). There was also an interaction (b = 1.76, SE = 0.81, p = 0.03). In the talk-and- navigate condition, pairs repeated story content more when the narrator was responsible for navigation than when the partner was. This was not the case in the navigate-only condition, possibly because no participant had an assigned role regarding the discussion (usually participants engaged in small talk while navigating in this condition, each contributing to the discussion).

Discussion

Talking together while walking together constitutes a complex set of concurrent joint activities. Using the example of turning, we have shown how the division of labor among pairs affects the coordination of the turn. Narrators used more verbal utterances to signal a turn than partners. This finding converges with those of Chevalley and Bangerter (2010) and Bangerter, Chevalley, and Derouwaux (2010), who found that it was more effortful for listeners to suspend a conversation than for speakers. In refraining from interrupting speakers, listeners also deployed more politeness, suggesting they were trying to mitigate the face threat of interrupting the speaker. In the present case, partners may have preferred to accomplish some signals via gesture, in order to avoid interrupting the narrator's story.

We also found that narrators repeated story-related utterances after a turn more often when they were navigators than when they were not, suggested that they were distracted by the double responsibility of narrating and signaling the turn. It may also be the case that this finding is related to the previous finding that narrators use more verbal means. Given that they have the floor, narrators may find it comparatively easier to interrupt their story to signal the turn. But in doing so, they may potentially interfere more with their own recall of where they were in the story than if they would use gestural means to signal the turn.

Our findings confirm that, in coordinating concurrent joint actions, participants need to manage common ground, pay attention to face wants of their partners, and that they may accomplish these constraints via a division of labor and using multimodal communication. Thus, coordinating concurrent joint actions expands the phenomenon of multitasking into the realm of conversational interaction and requires consideration of social as well as cognitive processes.

Acknowledgments

Research was supported by the Swiss National Science Foundation (Grant no. 100014_138195/1 to Adrian Bangerter). We thank Besarta Domaqi, Rosalie Muriset, Phillip Strahm and Lucie Voillat for assistance in conducting the study.

References

Bangerter, A., Chevalley, E., & Derouwaux, S. (2010). Managing third-party interruptions in conversations: Effects of duration and conversational role. *Journal of Language and Social Psychology, 29*, 235-244.

Bardram, J., & Bossen, C. (2005). Mobility work: The spatial dimension of collaboration at a hospital. *Computer Supported Cooperative Work, 14*, 131-160.

Bavelas, J. B., Coates, L., & Johnson, T. (2000). Listeners as co-narrators. *Journal of Personality and Social Psychology, 79*, 941-952.

Bavelas, J. B., Coates, L., & Johnson, T. (2002). Listener responses as a collaborative process: The role of gaze. *Journal of Communication, 52*, 566-580.

Brown, P., & Levinson, S. (1987). *Politeness: Some universals in language use*. Cambridge: Cambridge University Press.

Chevalley, E., & Bangerter, A. (2010). Suspending and reinstating joint activities with dialogue. *Discourse Processes, 47*, 263-291.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Enfield, N. J. (2006). Social consequences of common ground. In N. J. Enfield & S. C. Levinson (eds.), *Roots of human sociality: Culture, cognition, and human interaction* (pp. 399-430). Oxford: Berg.

Fussell, S. R., Kiesler, S., Setlock, L. D., Scupelli, P., & Weisband, S. (2004). Effects of Instant Messaging on the management of multiple projects. *CHI 2004* (pp. 191-198). NY: ACM Press.

Hutchins, E. (1995). Cognition in the wild. Cambridge, MA: MIT Press.

Kemper, S., Herman, R, & Lian, C. (2003). The cost of doing two things at once for young and older adults: talking while walking, finger tapping, and ignoring speech or noise. *Psychology and Aging, 18*, 181-192.

Lindenberger, U., Marsiske, M., & Baltes, P. B. (2000). Memorizing while walking: Increase in dual-task costs from young adulthood to old age. *Psychology and Aging, 15*, 417-436.

Relieu, M. (2008). Mobilité et coordination sociale. Note sur l'analyse ethnométhodologique des déplacements en situation naturelle. *Actes des 8èmes Journées Francophones Extraction et Gestion des Connaissances,* Sophia Antipolis.

Sacks, H., Schegloff, E. A. & Jefferson, G. (1974). A simplest systematics for the organisation of turn-taking for conversation. *Language, 50*, 696–735.

Salvucci, D., & Taatgen, N. (2011). *The multitasking mind*. Oxford: Oxford University Press.

Stivers, T., & Sidnell, J. (2005). Introduction: Multimodal interaction. *Semiotica*, *156*, 1-20.

Yatani, N., & Truong, K. (2009). An evaluation of stylus-based text entry methods on handheld devices studied in different user mobility states. *Pervasive and Mobile Computing, 5*, 469-508.

Zivotofsky, A. Z., & Hausdorff, J. M. (2007). The sensory feedback mechanisms enabling couples to walk synchronously: an initial investigation. *Journal of NeuroEngineering and Rehabilitation, 4*, 28.