Title
The emergence of procedural conventions in dialogue

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 33(33)

ISSN
1069-7977

Author
Mills, Gregory

Publication Date
2011

Peer reviewed

The emergence of procedural conventions in dialogue

Gregory J. Mills (gjmills@stanford.edu)
Department of Psychology, Stanford University
450 Serra Mall, Stanford, CA 94305

Abstract

A key problem for models of dialogue is to explain how conventions are established and sustained. Existing accounts emphasize the importance of interaction, demonstrating how collaborative feedback leads to more systematized, stable and arbitrary referential conventions. However, co-ordination in dialogue requires both co-ordination of content and of process. To investigate procedural co-ordination we report a collaborative task which presents participants with the recurrent co-ordination problem of ordering their actions and utterances into a single coherent sequence. The results provide evidence of the development of group-specific procedural conventions, resulting in elliptical utterances whose communicative meaning is determined by their sequential location within the dialogue.

Keywords: Dialogue; conventions; co-ordination; routines; sequentiality; language games; conversation analysis; (mis)communication.

Introduction

One of Wittgenstein's (1969) most quoted examples of language use concerns a building supervisor who utters the single-word command "Slab!" to refer to a slab that another builder is to retrieve. Wittgenstein uses this example to demonstrate how words depend for their meaning on the particular language game in which they are situated. For example, one could imagine a different group of builders using "paving stone" to refer to the same object. Alternatively, one could imagine another group of builders who use "Slab!" as a confirmation (after placing the object in the correct location).

A fundamental question concerning language use is how conventions such as "Slab!" are established and sustained. In recent years there has been a growing realization of the importance of addressing this question experimentally by observing how groups of participants develop their own conventions.

In what follows we argue that the basic findings of this research point towards the importance of the communicative use of language in interaction. We then argue that one of the key properties of dialogue is that it is underpinned by procedural regularities that have so far escaped direct experimental investigation. We then describe an experiment designed to test the hypothesis that these procedural regularities become conventionalized.

The importance of interaction

One of the earliest findings (Krauss and Weinheimer, 1967) in studies of conventions is that interlocutors' descriptions of novel objects rapidly converge on a shared set of referring expressions that become more concise on successive use. Importantly, this referential contraction does not occur in monologue (Clark, 1996).

Interaction in dialogue also directly affects the semantics of referring expressions: In a series of maze game experiments, Garrod and Doherty (1994) found that interlocutors who interacted with many different participants from the same sub-community used more abstract and systematized schemas than participants who had only interacted with a single interlocutor. Further, when interlocutors from one sub-group interacted with another sub-group they resorted to less abstract schemas (Healey, 2004). Similarly, experiments conducted by Schwartz (1995) on problem-solving compared the representations used by individuals with those used by dyads; here too, the representations used for communication were more abstract and systematized.

The importance of interaction has also been underscored by similar findings in graphical communication. The "Music Drawing Task" (Healey, 2007) required participants to create novel graphical representations for referring to pieces of music: Participants developed more systematic, abstract and concise representations if they were able to give each other feedback and reuse portions of each other's representations (see also Garrod et al., 2007). In another set of maze game experiments conducted by Galantucci (2005), participants who were given the collaborative task of creating novel graphical symbols for referring to locations in a maze developed highly systematized conventions for referring to maze locations.

Interaction also plays a central role in how interlocutors distinguish individual actions from communicative actions: When boot-strapping an entirely novel communication system, participants developed dyad-specific task-extraneous behavior to signal communicative intent (Scott-Phillips et al., 2009).

Despite these studies' very different approaches, a common methodological choice is their study of referential conventions, whether pieces of music, objects, shapes or locations in a maze. Put simply, the common question being asked is how two (or more) speakers of a community come to use the same signs in the same context, ideally on subsequent turns, to refer to the same referents. The immediate question that emerges is: If dialogue and interaction are of key importance, what exactly is it that interaction is contributing - are these the only kinds of convention that emerge in interaction?

Procedural organisation in dialogue

Empirical investigations of dialogue in Conversation Analysis (CA) demonstrate how dialogue is underpinned by procedural regularities across turns by different speakers that frequently consist of *different* kinds of utterance[1]: Questions are ordinarily responded to with an answer, not with another question; offers and invitations are ordinarily followed by acceptances or declinations (Levinson, 1983). These *adjacency pairs* (Schegloff, 1986) consist of a first pair part and a second pair part that operate normatively: Production of the first part creates an expectation that the second half is accountably "due" (Heritage, 1984), leading any response to be interpreted as pertaining to the second half. This locally managed system of local sequential coherence between turns results in global coherence through the hierarchical interleaving of embedded sequences that resolve local problems through, e.g. clarification, elaboration and reformulation (Levinson, 1983).

However, CA's primary concern with individual transcripts of naturally occurring dialogue has led to analyses eschewing experimental manipulation to probe specific predictions (Schegloff, 1992). As a result, CA has typically treated procedural regularities as static phenomena, already shared and known to be shared by interlocutors, and hence has not led to any systematic investigation of how they might develop during conversation.

The basic question this paper addresses is whether these procedural regularities existing between turns become conventionalized during interaction.

Methods

To test whether procedural conventions develop in a speech community, we drew on the methodology developed by Garrod and Doherty (1994) and Healey (1997) of assigning participants to different sub-groups. In the convergence phase of the experiment, participants only interact with members of the same sub-group. In the second phase, half the participants interact again with members of the same sub-group, while the other half interacts with members of the other group. This contrast between Within- and Cross-group communication allows a direct test of the development of group-specific conventions.

Crucially, the interlocutors are unaware of their assignment into different groups: Participants played a collaborative computer game which requires them to communicate using a text-based chat tool. All messages are passed through a server that obscures participants' identities, replacing them with automatically generated names that are changed on each trial.

[1] Cf. Millikan (2005) who proposes an alternative account of conventionalization (counterpart reproduction).

The alphabetical sorting task

In contrast to existing dialogue tasks that focus primarily on referential communication, the alphabetical task is designed to make reference as transparent as possible (the words are the referents), while presenting participants with the recurrent procedural coordination problem of ordering their utterances and actions into a single coherent sequence.

The task involves pairs of participants communicating with each other via a text-based chat tool (Healey and Mills 2006; 2011). Each participant's computer displays two windows: (1) A chat tool program used to communicate with the other participant; (2) A task window displaying the participant's score and a list of randomly generated words.

Solving the task requires participants to combine their lists of words into a single alphabetically ordered list. To ensure collaboration, participants can only select words that the other participant sees on their screen, and vice versa. To select a word, participants type a backslash "/", followed by the word. A further asymmetry introduced to ensure collaboration is that participants cannot see each other's selection, and if a participant makes a mistake, for example, by selecting an item that has already been selected, the *other* participant receives a status message saying that the item has already been selected. Participants are able to type "/restart" at any point to begin again.

For example, suppose Participant1 has the words BETA and GAMMA displayed on his screen, and Participant2 has the words ALPHA and OMEGA displayed on her screen: to solve the task, the participants need to communicate the lists of words to each other and then type the following:

(1)	**Participant1:** /ALPHA	
(2)		**Participant2:** /BETA
(3)		**Participant2:** /GAMMA
(4)	**Participant1:** /OMEGA	

Note that for an individual participant, ordering words alphabetically is a relatively straightforward task. However, for pairs of participants, this task presents the coordination problem of interleaving their selections correctly. Each set of words has only a single correct solution.

To hinder the development of group-specific referential expressions, different sets of randomly generated words were used in both phases.

Group assignment

24 participants were assigned to 6 sub-groups comprising 4 participants each. At any given moment, there are 12 conversations occurring simultaneously, relayed by the experimental chat tool.

The experiment was divided into two phases: (1) A convergence phase comprising 6 trials and lasting 40 minutes and (2) a test phase comprising a single trial lasting 5 minutes. To assist co-ordination, initial trials were slightly longer (see Table 1 below). On each trial participants see a new artificially generated name identifying their interlocutor, leading participants to believe they are speaking with a new partner on each trial.

Trial No.	Length
1	9 mins
2	8 mins
3	7 mins
4	6 mins
5	5 mins
6	5 mins
7 Test phase	5 mins

Table 1: Trial length (mins)

Convergence phase: Trials 1, 2, 3, 4, 5, 6.

In the first phase, participants alternated between speaking to two participants from the same sub-group. Every second trial, each participant was assigned to the same participant (alternating between the configuration of Figures 1 and 2 below).

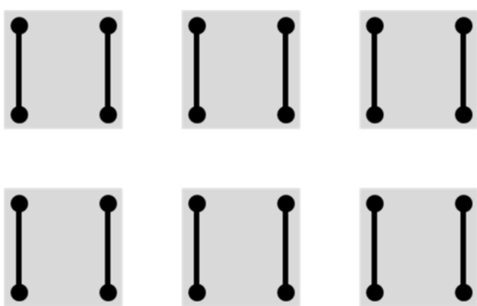

Figure 1: Trials 1, 3, 5. Circular dots represent participants and vertical bars represent participants interacting with each other.

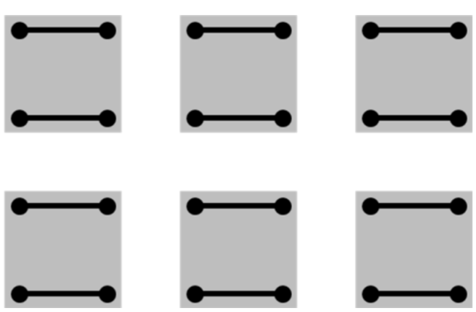

Figure 2: Trials 2, 4, 6. Circular dots represent participants and vertical bars represent participants interacting with each other.

Test phase: Trial 7

The second phase comprised a single trial, in which half the participants interacted with the remaining member of their sub-group (Within-Group), and the other half interacted with a participant from a different group (Cross-group).

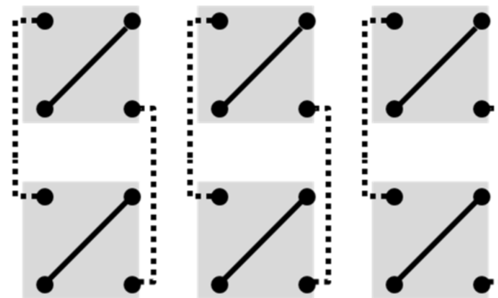

Figure 3: Trial 7. Black diagonal lines represent conversations *within* the same sub-group. Dotted vertical lines represent conversations *between* different sub-groups.

This second phase permits a focused comparison that is sensitive to the development of group-specific conventions: On the seventh trial all the participants interact with a participant with whom they have not interacted previously. However, participants in the Within-group condition are interacting with a participant who has interacted with both of their prior interlocutors. By contrast, those in the Cross-group condition are interacting with a participant from an entirely different sub-group.

Hypotheses

If the procedural conventionalization hypothesis is correct, Cross-group dyads will comprise participants who have developed different, sub-group specific procedural conventions. This should lead to participants in the Cross-group condition experiencing greater difficulty in co-ordinating their turns, and worse task performance than Within-group dyads.

Conversely, if the null hypothesis is correct, either there should be no difference between Within-group and Cross-group co-ordination, or members of the Cross-group should perform better, owing to the interlocutors benefitting from cumulative experience of two different sub-groups. (i.e. if there are no procedural differences, then all things being equal, the knowledge accrued from two groups' solution of problems should lead to better task-performance).

Results

24 Participants were recruited from the undergraduate student population of Stanford University, and received either course credit or payment ($15) for participating. On debriefing, no participant reported having detected that they were speaking with the same participant.

Overall, interlocutors produced 8246 turns, 794 turns of which were produced in the final phase.

Global development of co-ordination

To verify that participants become more co-ordinated at the task, the chat logs were used to determine the proportion of correct/incorrect solutions for each participant. Participants' performance in the first half of the convergence phase (trial 1, 2, 3) was compared with their performance in the second half (trial 4, 5, 6). Logistic regression showed a significant increase (χ^2 (1) = 11.1, p=0.01) from 58% correct to 75% correct. (see Figure 4, right).

Within vs. Cross-Group

Focusing on trial 7 (Test phase), to test for the effect of group-specific conventions on task performance, the proportion of correct/incorrect answers in Within-group dialogue was compared with Cross-group dialogue. Logistic regression showed Within-group participants generating significantly more correct answers (88%) than Cross-group participants (51%) (χ^2 (1) = 8.03, p<0.005).

To check that these differences are not due to individual participant differences, the same comparison was made between the same participants in the second half of the convergence phase (trial 4, 5, 6). Logistic regression yielded no differences between the pairs (χ^2 (1) = 0.07, p=0.79). Cross-group participants scored 73% correct answers, while the Within-group participants scored 75%, suggesting that the observed differences in trial 7 are due to the Within / Cross group manipulation.

Turn length

Focusing on the turns produced by participants, all things being equal, if different conventions make communication more difficult, then Cross-group participants should expend more effort in the Test phase in order to maintain co-ordination, and consequently produce longer turns than in Within-Group dialogue. A 1-way ANOVA yielded no effect F (1, 435) = 0.41, p = 0.52). Mean turn length = 14 chars.

Self-edits

To test for the effect of the intervention on participants' confidence, all participants' turns and selections were coded for whether they were edited prior to sending. Logistic regression showed participants in Cross-group dialogue editing their turns more (34%) than participants in Within-group dialogue (26%). (χ^2 (1) = 4.3, p = 0.036).

Typing speed

As a further measure of participants' confidence, keystroke data was used to determine the typing speed. All things being equal, if communication is made more difficult by different conventions, interlocutors should be more hesitant and type slower. The opposite is the case: 1-way ANOVA showed Cross-group participants typing faster (4.7 chars/sec) than Within-group participants (3.8 chars/sec).
F (1,435) = 9.63, p = 0.02).

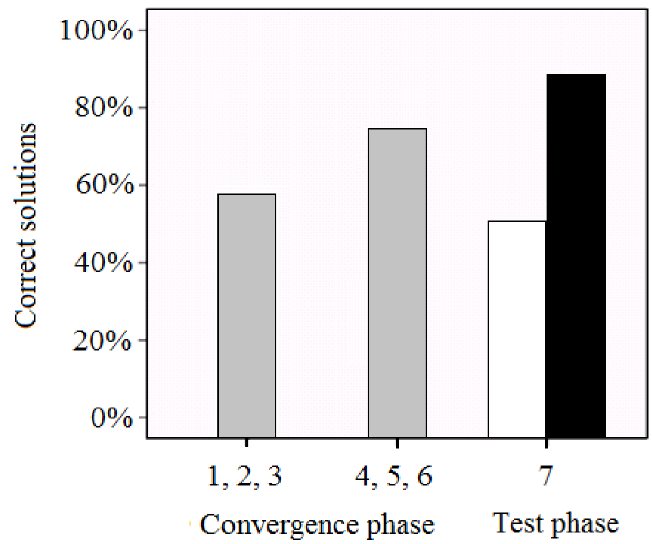

Figure 4: Proportion of correct solutions (White = Cross-group, Black = Within-group, Grey = Convergence phase).

Acknowledgments

Finally, all participants' turns were automatically classified according to whether they included acknowledgments such as "ok", "okay", "okey", "'kay". These acknowledgments have been shown to be deployed by interlocutors to explicitly demarcate boundaries between different stages of a dialogue (Clark and Bangerter 2004), for example using "ok, so" to prefix the start of a new topic.

If the procedural convention hypothesis is correct, Cross-group participants should experience greater difficulty in coordinating their transition from one section of the activity to the next, requiring the use of more acknowledgments to signal the boundaries between actions. This was confirmed: Cross-group dialogue contained more acknowledgments (12% of all turns), whereas only 5.6% of Within-group turns contained an acknowledgment. (χ^2 (1) = 4.4, p=0.035).

Discussion

The data provide strong support for the procedural convention hypothesis. Overall, participants become increasingly co-ordinated at the task. Participants in the Cross-group condition performed worse, editing their turns more, apparently spending more effort demarcating the boundaries in the dialogue. These observations might account for the faster typing speeds observed in Cross-group dialogue: If the participants are using different conventions for organizing their turns, this would place participants under greater pressure to time their own contributions correctly, hence the faster typing speeds, possibly exacerbating the problem and leading to the greater prevalence of typing errors.

Why is this finding important? Superficially it might appear unsurprising that groups develop their own conventions. However, the words being referred to are referentially transparent, and the task only permits a single solution that is trivially simple for an individual – what, then, are these group-specific properties that are causing interference in the Cross-group condition?

In the remainder of the discussion, we demonstrate that participants have developed procedural conventions for managing sequential constraints between turns, providing evidence that these conventions are not simply regularities, but are stronger normative conventions (Lewis, 1969).

(1)	Participant9:	Hold on. Ok?
(2)	Participant4	Done
(3)	Participant4	/AAAA
(4)	Participant9:	Have you selected anything?
(5)	Participant4	Yeah what did you select?
(6)	Participant9:	I'm done I did CCCC
(7)	Participant4	But I got to do BBBC first
(8)	Participant9:	Thought you said you're done?
(9)	Participant4	No I said you should
(10)	Participant9:	So which one have we selected
(11)	Participant9	So tell me what to choose now
(12)	Participant4	AAAAHH
(13)	Participant9:	Let's start again?
(14)	Participant4	Ok, tell me when you're ok
(15)	Participant9:	**/restart**

Transcript 1: Task dialogue from 2nd trial.

Conventionalizing idiosyncratic solutions.

The alphabet task requires participants to coordinate their turns: They cannot select each other's words, words can't be selected twice, and the words need to be selected in the correct order. This presents participants with the recurrent co-ordination problem of signaling who should select which words, as well as signaling the initiation and completion of these selections. As shown in Transcript 1, in the initial trials, participants barely manage to co-ordinate their actions. Here the participants are attempting (unsuccessfully) to order three words: "AAAA", "BBBB" and "CCCC", before restarting. (This pair took another 44 turns to solve this set of 3 words).

Over the course of the experiment, participants become progressively more co-ordinated, developing subtly different idiosyncratic solutions to the sequential co-ordination problem presented by the task. These include: Proposing and ratifying the full sequence in advance; negotiating each juncture incrementally; signalling pre- or post- selection (see Transcripts 2 and 3 below); avoiding complex signalling by co-ordinating on timing intervals.

Towards the end of the experiment, this culminates in the most co-ordinated pairs developing highly elliptical solutions, as outlined below: Transcripts 2 and 3 show two groups' different solutions to the same set of words ("AAA", "BBB", "CCC", "DDD").

(1)	Participant2	AAA, CCC
(2)	Participant4	BBB
(3)	Participant4	/AAA
(4)	Participant4	AAA
(5)	Participant2	/BBB
(6)	Participant2:	BBB
(7)	Participant4	/CCC
(1)	Participant2	AAA, CCC

Transcript 2: Highly elliptical dialogue from 6th trial

(1)	Participant3	BBB
(2)	Participant7	AAA, CCC,
(3)	Participant3	AAA
(4)	Participant3	/AAA
(5)	Participant3	BBB
(6)	Participant7	/BBB
(7)	Participant7	CCC
(8)	Participant7	/CCC

Transcript 3: Highly elliptical dialogue from 6th trial

A striking feature of these highly coordinated sequences is that (in contrast to Transcript 1) there are no overt questions, acknowledgments, or indeed any explicit descriptions of "what the turn is doing". In Transcript 2, prior to the critical juncture in turn (5), Participant4 says "AAA" in (4) to signal completion of the prior selection. In this dialogue a single word means[2] ~"I have just selected this word, now you do your selection, and tell me yours when you're done".

By contrast, in Transcript 3, the participants' utterances function differently. In turn (3), Participant3 types "AAA", to mean ~"I'm about to do AAA". Here, a word by itself means either ~"I'm going to select this word" or "You select this word" (as in turn 5).

Note here that although both dyads accomplish the task in almost the same number of steps, using highly elliptical sequences, all of which are in alphabetical order, what the single word utterances are doing at each point differs between the groups – using Wittgenstein's "Slab!" language game, "Slab!" in one group means ~"I selected the slab, now it's your turn", whereas in the other group it means ~"Now you select the slab".

Although each set of words only permits a single logical solution, it is striking to see how each group develops highly idiosyncratic procedural solutions. Focusing on the development of these elliptical sequences shows that in the initial trials, participants' attempts to explicitly negotiate a system more often than not prove unsuccessful. This finding resonates strongly with Pickering and Garrod (2004) who found in a series of maze game experiments that explicit negotiation was only successful once interlocutors had developed sufficient co-ordination. It appears that similar patterns occur in attempts by interlocutors to explicitly

[2] These explications of "what the turn means" are approximations for expository purposes.

negotiate procedures, as terms such as "now" or "wait", are inherently ambiguous: e.g. How long after uttering "now" still counts as "now"?; If the other participant says "ok", is that an instruction to stop waiting?

Patterns of (mis)communication

Focusing now on trial 7, when participant2 and participant3 subsequently interact with each other in the Cross-group condition, they encounter each other's different procedural uses of exactly the same words resulting in the following attempt to resolve the problem elliptically:

(1)	Participant3	/APPLE
(2)	Participant3	BAR
(3)	Participant2	BAR?
(4)	Participant3	Yeah of course
(5)	Participant2	WHAT?

Transcript 5: Misunderstanding in trial 7.

Here Participant3 says "BAR" to mean ~"Now you select BAR". However, Participant2 perceives Participant3 as having said ~"I've just selected BAR", which, to Participant2 makes no sense as it is neither the first word alphabetically, nor is it an object that Participant3 ought to select. Participant2's "BAR?" seeks clarification, but Participant3 perceives Participant2 as having asked ~"Do you want me to select BAR?", which Participant3 confirms, thereby perpetuating the misunderstanding.

Importantly, the prevalence of interlocutors initiating clarification requests, as in transcript 5, that query the procedural function of the elliptical utterances, indicate they are not simply regularities, but are actual conventions in the stronger normative sense: Participants notice deviations from what they expect to encounter and query their interlocutor for an account of what is missing.

Conclusions

This is to our knowledge the first experiment that has directly addressed the emergence of procedural conventions in dialogue. The results strongly suggest that they develop rapidly during interaction, over timeframes similar to those observed for referential conventions[3], leading to sequences of interleaved actions and utterances whose function is determined by their position within the sequence.

Acknowledgments

This work is supported by an individual Marie Curie IOF Fellowship: PIOF-GA-2009-236632-ERIS. I am grateful to Herb Clark and the anonymous reviewers of this paper for valuable comments and suggestions.

[3] One important caveat: It is currently unclear to what extent participants are associating these conventions with their interlocutors, as the manipulation of this study obscured the identity of their partners, making "strangers" seem more co-ordinated on each subsequent trial.

References

Bangerter, A., & Clark, H. H. (2003). Navigating joint projects with dialogue. *Cognitive Science*, 27, 195-225.

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

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive Science: A Multidisciplinary Journal*, 29(5), 737–767.

Garrod, S. & Doherty, G. (1994). Conversation, coordination and convention.. *Cognition*, 53, 181–215.

Garrod, S., Fay, N., Lee, J., Oberlander, J., & MacLeod, T. (2007). Foundations of Representation: Where Might Graphical Symbol Systems Come From? *Cognitive Science: A Multidisciplinary Journal*, 31(6), 961–987.

Healey, P.G.T. (1997). "Expertise or expert-ese: The emergence of task-oriented sub-languages." Proceedings of the 19th Annual Conference of The Cognitive Science Society. Stanford University, CA.

Healey, P. G. T. (2004). Dialogue in the degenerate case? Peer commentary on Pickering & Garrod (2004). *Behavioural and Brain Sciences*, 27(2), 201.

Healey, P. G. T. & Mills, G. (2006). Participation, precedence and co-ordination. In Proceedings of the 28th Conference of the Cognitive Science Society, Canada.

Healey, P. G. T. and Mills, G. (2011 in prep) DiET: The Dialogue Experimental Toolkit.

Healey, P. G. T., Swoboda, N., Umata, I., & King, J. (2007). Graphical Language Games: Interactional constraints on representational form. *Cognitive Science: A Multidisciplinary Journal*, 31(2), 285–309.

Heritage, P. (1984). *Garfinkel and Ethnomethodology*. Cambridge: Polity Press.

Krauss, R. M. and Weinheimer, S. (1966). Concurrent feedback, confirmation and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, 4 (3), 343–346.

Levinson, S. C. (1983). *Pragmatics*. CUP, Cambridge.

Lewis, D. (1969). *Convention: A philosophical study*. Harvard University Press.

Millikan, R. (2005) *Language: A Biological Model*, OUP.

Pickering, M. J. and Garrod, S. (2004). Towards a mechanistic psychology of dialogue. *Behavioural and Brain Sciences*, 27(2), 169–190.

Schegloff, E.A. (1986). The routine as achievement. *Human Studies*, 9 (2/3), 111-151.

Schegloff E. A. (1992) Repair after next turn *AJS* 97(5).

Schwartz, D. L. (1995). The emergence of abstract representations in dyad problem solving. *Journal of the Learning Sciences*, 4(4), 321–354.

Scott-Phillips, T., Kirby, S., & Ritchie, G. (2009). Signalling signalhood and the emergence of communication. *Cognition*, 113(2), 226–33.

Wittgenstein, L. (1969). *Philosophical Investigations*. Blackwell, Oxford.