A Dialogue Experimentation Toolkit

Gregory J. Mills
Stanford University,
Queen Mary, University of London

Patrick G. T. Healey
Queen Mary, University of London

Author Note

Gregory J. Mills. Dept of Psychology, Stanford University. Cognitive Science Research Group, School of Electronic Engineering and Computer Science, Queen Mary University of London; Patrick G. T. Healey, Cognitive Science Research Group, School of Electronic Engineering and Computer Science, Queen Mary University of London.

Gregory J. Mills is now at the Institute for Language, Cognition and Computation, University of Edinburgh.

This research was supported by grants from the Engineering and Physical Sciences Research Council (EPSRC/D057426/1-DIET) and by a Marie Curie Fellowship (PIOF-GA-2009-236632-ERIS).

Correspondence should be addressed to Gregory J. Mills, Institute for Language, Cognition and Computation, University of Edinburgh, EH8 9AB. gmills@staffmail.ed.ac.uk

Abstract

Dialogue is the native habitat for language acquisition and use. However, dialogue is also context dependent, fragmented and open-ended. This makes controlled experimentation difficult. Empirical research on dialogue has typically relied either on corpus analysis or on relatively coarse-grained experimental manipulations in which e.g. a confederate takes the role of one participant or in which joint tasks or channels of communication are configured to restrict interaction. Recent advances in natural language processing and the widespread use of text-based chat tools provide new opportunities for controlled experimental manipulations of live interaction. This paper presents 'DiET'; a text-based chat-tool platform developed to support fine-grained, word and turn-level interventions in free dialogue. This tool, which is distributed free to the research community, provides new ways to investigate the structural, procedural and conceptual organisation of dialogue. We explain the architecture and operation of the tool, the variety of interventions it can create, the kinds of interactional data it can capture and provide examples of its application to some long-standing problems in dialogue research.

Keywords: Dialogue, psycholinguistics, production, coordination, cooperation, social cognition

A Dialogue Experimentation Toolkit

Introduction

Language is first encountered, acquired and deployed in dialogue. Consequently, dialogue is a fundamental source of basic linguistic and psycholinguistic phenomena, and an important source of constraints on language structure and processing. This point can be illustrated by considering one of the most salient aspects of language use in dialogue; turn-taking (Sacks et al., 1974). Turn-taking is a basic feature of human language use in conversation. It has a foundational role in language acquisition (e.g. Beebe et al., 1988; Murray & Trevarthen, 1986; Scaife & Bruner, 1975) and appears to be a robust, cross-cultural universal (Levinson, 1983; Stivers et al., 2009).

The organisation of dialogue into sequences of turns that are produced by different interlocutors facilitates the use of specialised constructions. For example, fragmentary clarification requests such as “Who did?”, “Did he?” or “What?” incorporate constituents from other turns, produced by other interlocutors, as part of their meaning (Ginzburg & Cooper, 2004). Such highly context dependent turns are prevalent in dialogue where around 30% of utterances lack an overt predicate (Ginzburg, 2011). Clarification requests are themselves part of a wider class of dialogue-specific repair phenomena – the processes by which interlocutors signal and recover from problems with mutual-intelligibility. Repair processes are ubiquitous in dialogue and depend intrinsically on the structure of turn sequences to achieve their specific effects (Clark, 1996; Schegloff, 1987, 1992).

The internal structure of turns in dialogue is also determined, in part, by interaction. Turns have an incremental structure that is not predictable from syntactic or semantic units alone but also depends on patterns of concurrent feedback from other interlocutors (Clark & Wilkes-Gibbs, 1986; Goodwin, 1979; Howes et al., 2011). The incremental, interactional structure of turns is apparent in compound contributions (also termed ‘split turns’ or ‘co-constructions’) which consist of a single turn that is produced by multiple interlocutors (Howes et al., 2011; Lerner, 2002; Pickering & Garrod, 2013). Arguably, the clearest evidence for incremental processes in both language production and comprehension comes from the demands of interactive conversation (Brown-Schmidt & Tanenhaus, 2008; Pickering & Garrod, 2013).

Turn-taking is thus fundamental to human language use and involves fine-grained linguistic and temporal co-ordination (Gregoromichelaki et al., 2011; Lerner & Takagi, 1999; Poesio & Rieser, 2010; Schegloff, 2007). However, turn-taking and other dialogue phenomena are often not encountered in experimental studies of language processing. With some notable exceptions, discussed below, psycholinguistic studies have typcially focussed on investigating comprehension or production by single individuals in non-conversational contexts. The typical materials in psycholinguistic experiments are isolated words (e.g. Meyer et al., 1971), sentences (e.g. Bock, 1986; Magnuson et al., 2007; Swinney, 1979) or monologues (Bartlett, 1932; Brunyé et al., 2009; Thorndyke, 1977).

One risk of focusing separately on individual production and comprehension is that phenomena that appear as 'noise' or 'performance errors' may, in a naturalistic dialogue context, function as communicative signals (Clark, 1997). For example, disfluencies, hesitations, reformulations, repetitions and corrections that suggest processing problems for an indvidual can also play a constructive role in establishing and maintaining shared understanding in conversation (Arnold et al., 2003; Clark, 1996; Sacks, 1995; Schegloff, 2007; Schober, 1993; Schober & Clark, 1989). Hesitation markers such as "um" and "uhh" (Levinson, 1983; Clark & Fox Tree, 2002) are often used strategically to signal difficulty in granting another's request or as an invitation to others to speak (Schegloff, 2010). Ostensibly empty interjections such as "Oh", can, in context, indicate an important change of information state and are also used strategically to indicate a re-interpretation of a preceding turn (Heritage, 2002). Verbatim repetitions by an isolated speaker might appear to have a simple, informationally redundant function, However, verbatim repetitions are routinely used by conversational partners to produce humour, irony, expansion, elaboration, to demonstrate active participation, to confirm allusions and to perform repairs (Chouinard

& Clark, 2003; Jefferson, 1987; Heritage & Raymond, 2005; Schegloff, 1996; Tannen, 2007).

Perhaps the primary reason for the relative lack of experimental work on dialogue is the practical difficulty of carrying out controlled experiments (Pickering & Garrod, 2004; Brennan & Hanna, 2009). As the phenomenon of turn-taking highlights, dialogue is inherently interactive, open-ended, fragmented and context sensitive. This makes it difficult to achieve the kinds of experimental control that characterise psycholinguistic work on language processing in single individuals. Many contemporary psycholinguistic techniques are also difficult to apply to dialogue because they involve concurrent metalinguistic judgments, tasks and data collection procedures, such as fMRI and EEG monitoring, that either distort conversation or prevent it altogether (Scott et al., 2009; Tanenhaus & Brown-Schmidt, 2008).

Experimenting with dialogue

Although dialogue presents complex practical challenges, a number of useful experimental paradigms have been developed. These approaches typically involve experimental interventions that manipulate dialogue by placing general constraints on when and how people can communicate. We provide a brief overview of the main features and limitations of these approaches. Our argument is that, although productive, existing approaches are not fine-grained enough to unpick the mechanisms that underpin dialogue phenomena such as turn-taking and repair. We then present a chat-tool method that addresses some of these limitations by enabling experiments on live dialogue, including specific manipulations of clarification questions, compound contributions and cues to shared knowledge.

Manipulating Communication Networks

Some of the earliest experiments on dialogue investigated the effect of different patterns of interconnection between people on communciation i.e., manipulating who can interact with whom (e.g. Bartlett, 1932; A. Bavelas, 1950).

This work led to an increasing interest in modelling the topology of social networks (e.g. Freeman, 1979) and, in turn, further experimental investigations of how different network parameters, such as degree of connection and clustering, can affect interaction (e.g. Cassar, 2007; Freeman et al., 1980).

Experimental manipulations of opportunities for people to interact have been shown to have effects on both task outcome and dialogue processes; ranging from changes in the number and length of messages exchanged (e.g. Freeman, 1979) to the emergence of new group sub-languages or 'dialects' (Fay et al., 2000; Garrod & Doherty, 1994; Healey, 1997; Healey et al., 2007; Mills, 2011).

The popularity of computer mediated communication has enabled new applications of these ideas although primarily as an opportunity for large scale corpus analysis rather than controlled experiment (Garton et al., 2006; Kumar et al., 2010).

One limitation of this work is that models of network topology typically treat interactions as generic 'points of contact' that enable transmission of messages but abstract away from the ways in which the process of interaction itself (i.e. feedback, elaboration, clarification, hesitations, disfluencies, interjections) affects how a message is produced and comprehended (e.g. Bartlett, 1932; Kirby et al., 2008; Marin & Wellman, 2011) (although see Macura & Ginzburg, 2008 for a more complex model).

Manipulating Communication Channels

A second important strand of experimental work manipulates not only who can interact with whom but also the particular properties of the channels people use to communicate. One productive line of research has focused on manipulating the level of visual contact people have with each other during a dialogue. For example, through separation by a screen in co-present interaction (Boyle et al., 1994) or using video conferencing technology to enable different configurations of video and audio channels (e.g. Doherty-Sneddon et al., 1997; Olson et al., 1997; Whittaker, 2003).

The increasing sophistication of communication technologies makes more specific experimental manipulations possible. Early work looked at the effects of full verses half duplex audio on video mediated communication (O'Conaill et al., 1993) and subsequent studies investigated the effects of manipulating video parameters such as frame rate, bandwidth, and delay (e.g. Jackson et al., 2000; O'Malley et al., 1996; Whittaker, 2003). These systems can also be used to introduce asymmetrical constraints on communication channels using, for example, a bidirectional audio channel but unidirectional video (Kraut et al., 2003).

There has also been an increase in the sophistication of the dependent measures that can be obtained. The use of eye-tracking has allowed much more detailed analysis of the time course of linguistic processing in dialogue, as well as of the communicative function of mutual gaze (Richardson et al., 2008). Integration of real-time gaze-tracking with mediated communication allows participants' workspaces to be augmented with participants' gaze direction (Bard et al., 2007) and object fixation (Carletta et al., 2010).

More recent work, closer to the approach described in this paper, has exploited the targeted introduction of "noise" into an audio channel to induce mishearings (Schlangen & Fernandez, 2007) and the insertion of pauses and hesitations into turns as specific signals of speaker difficulty (Arnold et al., 2007; Brennan & Schober, 2001). Comparisons of computer-mediated text messaging tools have also been used to explore the effects of different turn-taking mechanisms on interaction (Herring, 1999; Phillips, 2007).

A general finding emerging from this work is that people compensate for restrictions in the communication channels available to them. For example, when people cannot see each other during a task-oriented dialogue they compensate by almost doubling the number of turns and words they use in order to preserve task performance (Doherty-Sneddon et al., 1997). This highlights how manipulations of system level parameters such as frame-rate and delay are generally too coarse-grained to unpick the underlying processes that enable successful interaction using these systems (Whittaker, 2003). However, once communication is technologically mediated, more sensitive real-time processing of some communicative signals becomes possible. This technique, exploited by the DiET tool described below, makes it possible to detect and transform a behaviour before it is presented to a conversational partner. This approach has been used to create artificial representations of interlocutors' gestures (Gratch et al., 2006), facial orientation (Bailenson et al., 2008) and gaze and head pose in turn-taking (Edlund & Beskow, 2009).

Task Manipulations

Perhaps the most extensively used experimental technique in studies of dialogue is manipulation of a shared task. One reason for this is the useful measures of outcome, such as quality, accuracy and completion time that a task can provide. The use of target stimuli and objects also helps to simplify conversation by providing a relatively simple, extensional index of what people are talking about (e.g. (Anderson et al., 1991; Brennan & Clark, 1996; Garrod & Anderson, 1987; Krauss & Weinheimer, 1967; Metzing & Brennan, 2003; Wilkes-Gibbs & Clark, 1992)).

Task manipulations have been used to explore a variety of hypotheses. Some early studies varied the familiarity of task materials in order to examine their effects on e.g. ease of story transmission (Bartlett, 1932) and the difficulty of referring successfully (Krauss & Weinheimer, 1967). Confusability of target items has been used to investigate factors influencing the construction of referring experssions (Brennan & Clark, 1996; Wilkes-Gibbs & Clark, 1992). Studies of the Map task introduced systematic mismatches in what each participant knows (Anderson et al., 1991), a strategy also used in work that manipulates differences in who can see what (Barr & Keysar, 2002; Clark & Krych, 2004; Hanna & Brennan, 2007). Task materials and communication media have also been developed that deliberately impede people's ability to use existing communicative conventions, forcing them to develop new 'languages' on the fly (Galantucci, 2005; Healey et al., 2002; Kirby et al., 2008).

A number of studies have used task manipulations to control people's degree of involvement in an interaction. People's basic level of engagement and responsivity in a dialogue can be manipulated using concurrent 'distractor' tasks (J. Bavelas et al., 2002; J. Bavelas & Gerwing, 2011). Different kinds of participation such as 'side-participant', 'bystander' and 'overhearer' can be manipulated through task assigned roles and affect levels of shared and mutual knowledge (Barr & Keysar, 2002; Branigan et al., 2007; Schober & Clark, 1989; Wilkes-Gibbs & Clark, 1992). A further development of this technique involving switching task roles has been successfully used to explore how people maintain or abandon locally negotiated referring expressions (Brennan & Clark, 1996; Metzing & Brennan, 2003).

Although task manipulations are one of the most productive paradigms in experimental research on dialogue, they are typically used to introduce constraints that set parameters for the interaction as a whole, rather than selectively targeting particular dialogue mechanisms (although see for example J. Bavelas & Gerwing, 2011).

Confederate Techniques

Perhaps the most direct form of experimental control over word and turn-level dialogue phenomena has been achieved by the use of confederates who make scripted or partially scripted contributions to interaction. An early example of this approach is Garfinkel's (1967) 'breaching' experiments, in which confederates were instructed to act out particular roles, such as pretending to use specific styles of questioning. Experimental tests of communicative signals of affiliation and empathy have made extensive use of confederate techniques to manipulate repetition of non-verbal signals such as gaze and gesture (J. Bavelas et al., 1986; Chartrand & Bargh, 1999) and repetition of another persons words (Van Baaren et al., 2003).

Confederate techniques have also been used to introduce experimental probe words into dialogue to test the conditions under which conversation partners will incorporate them in their own turns (Berko, 1958; Clark, 1979; Levelt & Kelter, 1982; Saxton, 1997). This technique has been used to explore the extent to which people will repeat each other's syntactic constructions in dialogue (Branigan et al., 2000, 2007). Confederates also provide a way to investigate the effects of third-party interruptions on speakers and hearers (Chevalley & Bangerter, 2010).

While confederate techniques offer the most direct way of targetting specific dialogue phenomena they are also one of the most problematic. Scripted prompts are inevitably less naturalistic and spontaneous than ordinary utterances (J. Bavelas et al., 2008; Beattie & Aboudan, 1994; Brennan et al., 2010). It is difficult for confederates to control their behaviour naturalistically while contending with the real-time demands of social interaction, and many dialogue mechanisms are not straightforwardly under conscious control (Pentland, 2004; Pickering & Garrod, 2004). If a confederate is not blind to the experimental intervention then they may also inadvertently cue naïve participants to a manipulation. Even where they are blind, the effect of repeating a task more often than other participants also alters the naturalness of their behaviour. One strategy for coping with these problems is to restrict confederates' non-target behaviours in order, for example, to avoid engaging in 'off-task' conversation, however this further limits the naturalness of the dialogue (Brennan et al., 2010; Kuhlen & Brennan, 2013). Training confederates to be more consistent across trials can have the undesired effect of them becoming "de facto" experts at the task. This can affect how confederates respond to participants' informational needs throughout the task, even in the unscripted parts of the interaction (Brennan & Hanna, 2009).

Part of the problem with confederate studies is the emergent nature of dialogue: many dialogue phenomena, such as referential contraction (Clark, 1996), alignment (Pickering & Garrod, 2004) and audience design (Brennan et al., 2010) only occur as a consequence of tacit, reciprocal adjustment by interlocutors to each other as the interaction unfolds.

Summary

Although there are a variety of well established techniques for experimenting with dialogue, the experimental interventions used have not achieved the fine-grained word and sentence level control found in psycholinguistic studies of individual language production and comprehension. The study of individual language processing benefits from established experimental paradigms (e.g. priming and cued recall), standardised measures of cognitive activity (e.g. gaze tracking and brain imaging) and a variety of experimental toolkits that allow for highly configurable, controlled and replicable experimental designs and data collection, such as PsychToolBox (Kleiner et al., 2007), E-Prime (Schneider et al., 2002), PsyScope (MacWhinney et al., 1993).

By contrast, studies on dialogue are faced with the problem of managing a steep trade-off between experimental control and validity: Currently, the sole technique that permits direct manipulation of both the content and the timing of participants' contributions, while retaining interactivity, requires the use of confederates. However, this is also one of the most problematic techniques.

The limitations on existing experimental methods have held back progress in the analysis of dialogue and the major challenges identified by Krauss and Fussell (1996) remain largely unaddressed, namely to go beyond the study of reference to concrete objects; to deal with multi-party conversations; to analyse the effects of communication (grounding) failures and their impact on (re)interpretation; to distinguish between the effects of collaborative and non-collaborative processes and to address individual and socio-cultural differences in communication.

DiET: A Dialogue Experimentation Toolkit

DiET addesses the limitations of current experimental techniques by combining a text-based chat tool with natural language processing techniques. This enables participants' contributions to a dialogue to be automatically intercepted, analysed, selectively transformed and relayed to other participants in real-time. This makes a variety of systematic, context-sensitive experimental manipulations of live, unscripted interaction become possible for the first time.

Importantly, the selective transformations of a dialogue turn enabled by DiET are seen only by other participants. This allows the introduction of experimental dialogue manipulations with minimal overt disruption to the dialogue and without participant's awareness.

We proceed by describing the architecture of the DiET toolkit, the basic set of client and server based interventions, the general scripting tools for creating a new intervention and the format of the data files provided by the distribution freely available from http://cogsci.eecs.qmul.ac.uk/diet/ under a General Public License (GPL). We discuss some of the practical considerations in using this approach to experimenting with live dialogue. We then summarise three sets of experiments that have explored the use of this technique to investigate the effects on dialogue processes of 'spoof' clarification questions, compound contributions and the grounding signals that establish common ground.

Architecture

The basic experimental setup involves two (or more) participants seated in separate rooms. Each participant sits in front of a computer which is connected via a network (cable or wireless) to the computer which is used by the experimenter. The chat tool is a client-server application, written in Java. It consists of the following two components:

- The **client software** runs on each participant's computer, either as a stand-alone program, or loaded from a webpage. It provides a configurable interface that displays the unfolding conversation. All the text typed by participants as well as additional data collected by the interface (see below) is recorded and passed to the server.

- The **server software** runs on a single computer. It acts as intermediary, relaying participants' turns, and also collates and stores all the participant data that is captured by the clients. It provides an interface for the experimenter to monitor the unfolding conversation. The main function of the server is to allow experimental control over the interaction: It provides a powerful set of scripting tools for creating interventions that intercept and manipulate the dialogue in real-time.

The two main kinds of intervention this permits are **(1) Client-based interventions** that manipulate the interaction by configuring the interactional constraints of the chat tool interface, and **(2) Server-based interventions** that manipulate what a participant perceives the other participant(s) as having typed, e.g. manipulating the content, timing or apparent identity of the participant who produced a turn.

Client-Based Interventions

The chat tool includes 3 configurable interfaces that support experimental control over how the unfolding dialogue is displayed on participants' screens. The interfaces also support interventions that place constraints on participants' ability to formulate turns. All the features described below can be configured in the experimental script.

Single window chat interface. This is the standard interface used in most proprietary chat programs (e.g. Skype or Facebook). The interface consists of a primary window that displays the unfolding conversation, a separate text-entry window, and a *status bar* that displays information about the typing activity of the other participants (see Figure 1). Participants formulate their text in the text entry window, which allows them to edit their turn privately prior to sending. Each new turn is displayed in the primary window, on a separate line, and is prefixed with the participant's name. This interface supports the experimental manipulation of:

- **Dialogue history:** Participants' access to the dialogue history can be controlled by restricting the size of the window, by inhibiting scrolling, or by controlling how many prior turns are visible to the participants.

- **Revisability:** Participants' ability to formulate turns can be manipulated by modifying the properties of the text-entry window (e.g. by restricting the number of characters, or by inhibiting participants' ability to edit their turns).

- **Status messages:** In proprietary chat clients, status messages are used to indicate typing activity. These messages can be selectively blocked or modified to make it appear as if another participant is typing. The status bar can also be modified to display task-specific status messages.

Multiple window interface. This interface displays participants' turns in separate windows. Figure 2 shows the dual window version of this interface which has an upper window that only displays participant A's turns, and a lower window that only displays turns produced by participant B.

This interface permits the experimental investigation of dialogue coherence, as it obscures the sequential ordering between participants' turns, while retaining their informational content. For example, in Figure 2, it is unclear whether A's second turn *"Ok is it this one"* is responding to B's first turn *"it's like the other one"* or to *"what shall I do?"*. Similarly, it is unclear whether B's third turn *"what?"* is asking for clarification of A's *"Ok is it this one"* or is attempting to clarify *"now place your shape there"*.

This interface also supports the experimental manipulation of dialogue history, revisability and status messages.

Character-by-character interface. An oft-cited criticism of text-based interaction is that participants' ability to formulate and revise their turns prior to sending decreases the coherence in the interaction, due to the potential for parallel conversational threads (Herring, 1999; Garcia & Jacobs, 1999; Phillips, 2007). The character-by-character interface suppresses this possibility by enforcing public, incremental turn-formulation of the kind used in the early 'talk' facility [1].

[1] Unix BSD v4.2, 2983

The interface consists of a single, main window which is used both for text-entry and for displaying the unfolding conversation. The window displays characters as they are typed, immediately on all participants' screens. In contrast with the preceding interfaces, participants formulate their turns in "public", allowing them to observe each other's incremental turn production and revision. By displaying all participants' contributions incrementally in a single window, this interface requires participants to co-ordinate their turn-taking behaviour with each other (Sacks et al., 1974), leading to dialogues that are more analogous to the sequential structure of spoken, conversational interaction.

This interface also supports the manipulation of dialogue history, revisability and status messages. Additionally, this interface provides experimental control over participants' ability to request, hold and relinquish the 'conversational floor' (Sacks et al., 1974), e.g. allowing control over who can interrupt whom.

Server-based Interventions

The main functionality of the chat tool is to support experimental scripting of interventions that selectively manipulate participants' turns. Processing by the server means that interventions can be made, in real-time, in ways that are sensitive to the unfolding conversational context. This typically involves specifying criteria for identifying target turns in the interaction, followed by selective transformation of the content of these turns (adding, deleting or replacing constituent elements). Scripting also allows selective blocking of specific turns, as well as the insertion of artificial turns into the dialogue. Additionally, the chat tool can be used to modify the apparent identity of the participant who produced the turn.

Crucially, participants are unaware of any experimental manipulation performed by the server, and, unlike confederate techniques, this technique allows the interventions to be made both fully "blind" and inserted into an unscripted exchange. In addition, because the server can filter who sees what, controlled asymmetries between multiple conversational participants can be created. Potential interventions include:

- The insertion, blocking and transformation of disfluencies, self-corrections and hesitations to investigate **inter-individual effects of production errors**.
- Systematic substitution of synonyms, hypernyms and hyponyms to investigate **conceptual co-ordination, alignment and entrainment**.
- Manipulation of apparent turn origin, turn sequencing, timing, and dialogue history to investigate the **procedural organisation of conversation**.
- Manipulation of semantic and syntactic constituents to investigate **co-ordination of linguistic structure**.
- Manipulation of specificity and type of feedback (acknowledgments, repair) to investigate signalling and resolution of **miscommunication**.
- Introducing, transforming or selectively blocking **discourse particles** to investigate their effect on **dialogue coherence**.
- Manipulation of participatory status, and turn origin to investigate **multi-party dialogue**.
- Assigning participants to different sub-groups to investigate the emergence of **group-specific conventions**.
- Manipulation of perceived **identity, gender, group-affiliation, levels of co-operation.**

Scripting Interventions

DiET provides a rich set of scripting tools (API) for specifying the target criteria of the interventions, and also for controlling how the interventions unfold. All scripting occurs on the server, in Java.

The chat tool is designed so that all information from the chat clients is sent instantly to the server. The script on the server functions as an intermediary that specifies what happens to this information — the default script simply relays the turns between all participants and also instructs the chat interfaces to display the appropriate *“Participant X is typing”* status messages. Scripting interventions involves customizing this behaviour to create targeted interventions, as illustrated below, in pseudo-code.

Normal operation. The default script that relays turns between all participants looks as follows: Here, on receiving a `Turn` from a participant called `Client`, the script simply instructs the server to send this turn to all the other clients.

```
processTurnFromClient(Turn, Client){
        sendTurnToAllOtherClients(Turn)
}
```

Transforming the content of a turn. The example below shows an intervention that inserts a disfluency marker “umm" (Clark & Fox Tree, 2002) at the beginning of a turn:

```
processTurnFromClient(Turn, Client){
        FakeTurn = ‘‘umm ’’ + Turn
        sendTurnToOtherClients(FakeTurn)
}
```

Inserting an artificial turn. The example below shows pseudo-code for scripting the first 2 steps of the intervention detailed in Figure 1. This intervention inserts artificial ‘fragment’ clarification requests (Purver et al., 2003) which query individual constituent words of a previous turn. The script inspects all participants’ turns, and on detecting a target turn that contains a suitable word, e.g. *“top”*, the chat tool generates a clarification request that appears to originate from another participant, and sends it to the producer of the target turn.

```
(1) processTurnFromClient(Turn, Client){
(2)     sendTurnToOtherClients(Turn)
(3)     FakeTurn = findClarification(Turn)
(4)     spoofReadingDelay(FakeOrigin, Turn)
(5)     spoofTyping(Client, FakeOrigin, FakeTurn)
(6)     sendTurn(FakeTurn, Client, FakeOrigin)
(7)  }
```

In this example, line (2) instructs the server to send the orignal `Turn` produced by `Client` to the other dialogue participants. Line (3) inspects the `Turn` in order to determine whether it contains a word that is suitable for clarifying. On detecting a suitable word, this is followed, in line (4), by an artificial delay that simulates `FakeOrigin` reading `Client's` turn. Line (5) instructs the chat tool to make it appear, to `Client` as if `FakeOrigin` is typing the response. This is accomplished by displaying an artificial status message *"FakeOrigin is typing"* on `Client's` interface. The duration of this status message is determined by using `FakeTurn`'s average typing time. After constructing the artificial clarification request *"sorry what?"* in line (5), in line (6) the server sends this clarification request to `Client`.

As written, this script would simply parrot back a clarification request for every turn. An actual implementation might randomly query a sub-set of all turns, or use task-specific target criteria to trigger and select the form of the intervention. Further scripting would be needed to ensure that the response to the clarification request isn't relayed to the other participants, and also to ensure that the response is in turn responded to with an acknowledgment (See Figure 1 and also the first example study detailed below). The library of scripts included with the chat tool contains templates that can be customized for this purpose.

Creating plausible, naturalistic interventions

The API includes resources for creating more plausible, naturalistic interventions. The data captured dynamically during the experiment allows the artificially generated turns to mimic the timing of individual participants' typing behaviour, and also allows the experimenter to introduce natural looking typing mistakes ("typos") into the dialogue. The dialogue history can also be used to create interventions that are sensitive to participants' idiosyncratic language use.

Experience with this approach has shown that it can be important to restrict the frequency of interventions. As (Garfinkel, 1967) noted, too many clarification questions can provoke irritation if not anger from other conversational participants as it can easily create the impression that someone is being deliberately uncooperative. This makes pilot testing of new interventions essential.

Data Collection and Storage

The chat tool produces output that is immediately available for statistical analysis, (obviating the need for costly and time-consuming transcription of participants' dialogue). For each turn, the chat tool records the origin (who typed the turn), the recipients (which participants actually received the turn), the apparent origin (who the participants were led to believe typed the turn), as well as the timestamp of each keypress. The chat tool also records the turn onset time, turn formulation time, typing speed, and typing overlap between participants' turns. As measures of turn-formulation difficulty, the chat tool records the number of self-edits (deletes, insertions) and their position within the edited text. In addition, the frequency and recency of each word used by the participants is also stored.

This data is available at runtime via the API, and can be used by the experimental scripts to generate interventions that are sensitive to the dialogue history and to the immediate local conversational context.

All data collected during the experiment is automatically saved to a CSV (Comma Separated Values) text file that can be easily loaded into statistical or spreadsheet software.

Participants

Participants are assigned participant IDs, allowing anonymisation of the chat tool logs and also allowing experiments to be run 'double-blind'.

Stimuli

The current distribution of the chat tool includes two tasks: (1) A variant of the maze game task (Garrod & Doherty, 1994; Healey & Mills, 2006); (2) A variant of the tangram task (Clark & Wilkes-Gibbs, 1986). The chat tool also includes scripts for customizing novel "visual world" referential tasks (Tanenhaus & Brown-Schmidt, 2008).

Stimuli can either be scripted directly in Java, or designed as webpages, in HTML. The presentation (e.g. ordering, persistence, timing) of the stimuli can be controlled experimentally by the server, and can be made contingent on properties of the dialogue. The chat tool can also be used to control (via networking) other experimental software, e.g. MATLAB experimental toolkits.

Participants' responses to all stimuli are saved to the log file, and are instantly available as parameters in the experimental script, permitting the design of highly context-sensitive interventions.

Remote participation and longitudinal studies

The chat client software can be run directly from a web-browser, without installing any additional software. This allows greater flexibility in conducting experiments, in particular longitudinal studies, as participants can participate from different locations (i.e. different laboratories or from their home computer). In addition, the chat tool includes libraries for interfacing directly with crowd-sourcing technologies (Munro et al., 2010) for recruitment and participant compensation.

Integration with other software

The chat tool includes resources for interfacing directly with the Stanford Parser (Klein & Manning, 2003) and WordNet (Miller, 1995) and other Java-based software. Integration with non-Java based NLP programs, in particular Python, can be achieved with Jython, or JPype.

Example Studies

The chat tool includes a constantly growing library of interventions that have been thoroughly tested and published. These interventions are available as templates that can be adapted for new experiments. Three of these interventions are detailed below:

Artificial Clarification Requests

An especially powerful technique uses artificially generated probe questions (e.g. *"top?"* in Figure 1) in order to engage participants in short clarification subdialogues. The basic idea behind this technique is that querying participants' turns provides a useful probe for naturalistically exploring the possible interpretations that interlocutors assign to each other's utterances.

This technique offers a high level of experimental control as interventions can be scripted to target specific turns that are of theoretical interest (e.g. targeting specific lexical, syntactic or semantic structures, as well as turns that occur at key moments in the dialogue task). Scripting also allows control over the specific form of the clarification request (e.g. *"top?"* vs. *"where exactly is the top?"* vs. *"where?"*).

Importantly, this approach minimizes the disruption to the rest of the dialogue, as unlike studies using confederates, once the clarification sub-dialogue is completed, the participants can continue with fully unconstrained dialogue. Studies using this technique include:

- Investigating the effect of clarifying first and second mention of content vs. function words (Healey et al., 2003).
- Inserting artificial clarification requests that signal different levels of ‘trouble’ to investigate their effect on semantic co-ordination (Mills & Healey, 2006).
- Interrupting participants’ turns at clausal boundaries to investigate incremental language production (Eshghi et al., 2010).
- Manipulating the apparent origin of clarification requests to investigate partner-specific indexing of levels of co-ordination (Healey et al., 2011).
- Clarifying participants’ discourse-plans (Mills, 2012a).
- Clarifying the procedural function of participants’ turns (Mills, 2012b).
- Capturing naturally occurring clarification requests and transforming them, e.g. by increasing their severity.

Engaging a participant in an artificial clarification subdialogue typically involves the following distinct stages, which correspond to the individual states in the flowchart in Figure 1).

1. **Detecting a target turn.** The experimenter will typically have written a script that specifies criteria for identifying an appropriate target turn, e.g. detecting the use of a particular word or syntactic construction. On detecting the target turn, the chat tool relays the turn to the other participants, before initiating the artificial scripted clarification sequence.
2. **Blocking the other participants**. The server instructs all chat client interfaces, other than that of the recipient of the clarification, to prevent participants from typing text in their turn formulation windows. The server also instructs these blocked clients to display an artificial error message *“Network Error...please wait”* on their status bar.

This ensures that while the recipient of the clarification request is responding, the other participants are prevented from typing until step 8 below, and are given a plausible reason for this prevention.

3. **Artificial indication of typing activity.** The server instructs the chat client of the recipient to display a "spoof" status message (e.g. *"Participant X is typing"*), which makes it appear as if the apparent producer of the clarification request is currently typing the clarification request.

4. **Sending the artificial clarification request**. The server consults the experimental script to determine the form of the clarification request (e.g. *"sorry what?"* or *"top?"*), and sends it to the recipient.

5. **Waiting for the response**. If the recipient takes longer than a pre-specified time, the server aborts the intervention, and re-enables all the other participants' interfaces.

6. **Intercepting the response**. On receiving the response (which is not relayed to the other participants), the server pauses for a brief duration in order to simulate the apparent producer reading the response. This duration is determined automatically by using the data collected during the experiment to calculate the average response time of the apparent producer.

7. **Acknowledging the response**. The server sends a "spoof" acknowledgment from the apparent origin, e.g. *"ok"* or *"ok thanks"*. As in step 3, this "spoofed" acknowledgment is also preceded by artificial status messages that mimic the typing behaviour of the apparent producer.

8. **Resume.** The server re-enables the other participants' interfaces, allowing the dialogue to resume.

Compound Contributions

Another set of experiments (Howes et al., 2011, 2012) investigated compound contributions (also referred to as 'split-utterances' and 'co-constructions') (Lerner, 2002; Lerner & Takagi, 1999; Poesio & Rieser, 2010). These are contributions that are produced by more than one speaker, in consecutive turns, for example:

(1)
A: I think we should go to the shop to
B: get some food

The experiment conducted by Howes et al (2011) investigated how compound contributions produced by two participants affect another participant's perception of the state of the dialogue . Four participants communicated with each other via the DiET chat tool, in a problem-solving task. The chat tool was used to intercept ordinary turns and selectively transform them into artificially generated compound contributions. Participants' responses to these completions were subsequently analyzed (see Figure 3).

The interventions were scripted as follows: On detecting a suitable target turn produced by one of the participants, for example:

(2) Participant1: I think that is a really great idea.

The chat server splits this target turn into two halves of an artificial compound contribution, yielding a first half *("I think that is")* and a second half *("a really great idea")*. The chat server generates 4 different variants that appear to have been produced jointly by *different* interlocutors; each recipient of the intervention perceives a different participant as having initiated and completed the compound contributions. For example, the target turn (2) would be transformed into the following variants:

Variant 1. The first half appears to be produced by the participant who originally typed the target turn, and the second half appears to be produced by a *different* participant. This variant is received by Participant 2:

(v1)

Participant1:	I think that is
Participant3:	a really great idea

Variant 2. The first half appears to be produced by a different participant, and the second half appears to be produced by the same participant. Received by Participant 3:

(v2)

Participant2:	I think that is
Participant1:	a really great idea

Variant 3. Both halves appear to be produced by different participants. Received by participant 4.

(v3)

Participant2:	I think that is
Participant3:	a really great idea

The interventions were accomplished in the following 7 steps:

1. **Identifying a target turn.** The server analyzes participants' turns in order to detect turns that are suitable for transformation. A blacklist is used to exclude problematic turns (e.g. turns that contain personal pronouns).

2. **Transforming the target turn.** The target turn is parsed to identify a phrase

boundary where it can be divided into the two halves of an artificial compound contribution. On detecting a suitable boundary, the chat tool generates the 3 variants of the compound contribution, as described above.

3. **Sending the first half.** The server sends the first half of each compound contribution to its respective recipient.

4. **Artificial turn-taking behaviour.** From the chat logs, the server calculates the average time that participants take to respond to each other, and pauses for that duration. This pause simulates the apparent producer of the second half processing the first half, prior to initiating a response.

5. **Artificial typing activity.** The server instructs the chat clients to display 'spoof' status messages (e.g. *"Participant X is typing")* which make it appear as if the apparent producer of the second half of the compound contribution is currently typing. The duration of these messages is calculated from participants' average typing speed.

6. **Sending the second half of the compound contribution.** The chat tool sends the second half, mimicking the apparent origin's typing behaviour.

7. **Resume.** The chat tool resumes normal operation, and the participants continue with naturalistic interaction.

Grounding Downgrades

Another set of experiments investigated how interlocutors in multi-party dialogue adapt to perceived difficulties of another interlocutor. Participants played a 3-player version of the maze game (Garrod & Doherty, 1994), involving one Director (D) and two Matchers (M1 and M2).

The chat tool was used to identify Matchers' turns that contained explicit acknowledgments, such as "ok". While the director received the original turn, the other

Matcher received a modified turn which replaced the acknowledgments with more tentative grounding cues, such as “hmm” or “well". This is illustrated below in Table 1.

Discussion

The combination of text-based chat with natural language processing techniques make a much wider range of dialogue phenomena open to direct experimental investigation, at a much higher resolution than has previously been possible. This has enabled some of the first selective experiments on key dialogue phenomena including; the effects of context and form on the processing of clarification questions, the ability to create and induce compound contributions and the effect of selective ‘downgrading’ of grounding cues on shared context.

The ability to introduce interventions into free dialogue also considerably expands the potential range of conversational contexts that can be explored and addresses problems with using confederates. An additional advantage of this technique is that it allows selective targeting of specific dialogue phenomena, while leaving the rest of the dialogue untouched. Judicious scripting of interventions permits robust interventions that are extremely difficult for participants to detect. To date we have conducted experiments on over 1200 participants. Analysis of the dialogue transcripts and debriefing interviews yielded 8 participants who detected the artificial interventions. In 6 of these cases, the participants were siblings or very close friends who interacted with each other on a daily basis, using text-based chat.

As noted in the introduction, some recent experiments on speech and gesture in dialogue have exploited a similar strategy of selecting transformation of communicative signals in real-time. An important advantage of text chat in this respect is that the timing of chat exchanges is such that a complete turn can be processed and transformed before a response is expected or produced. This is harder to achieve with other modalities. For example, substitution of synonyms in live speech requires that the target word is recognised prior to substitution, which introduces a noticeable delay in the transformed turn or, if buffering is used, in the exchange of turns. Similary, interventions with non-verbal signals, such as gestures or expressions, that are contingent on processing their content or form cannot currently be introduced without perceptible disruption to live interaction.

Part of the motivation for the chat-tool approach used in DiET is to improve the naturalism and validity of experimental research on dialogue. This begs the question of how natural text-chat is as a mode of communication. The answer to this question is relatively straightforward; text chat in various forms has been in use for over 20 years and is now a very widely used form of communication. The primary means of communication in social networking sites (e.g. Facebook, Twitter), and also in collaborative learning environments (e.g. Coursera) is via text-based interaction. The most popular proprietary text-messaging applications currently (2013) have over 300 million users each, and handle over 10 billion messages per day (Bradshaw, 2013). In 2012, American teenagers sent a median of 60 sms text messages per day (Lenhardt, 2012).

It also begs the question of how well the results from such studies can generalise to other modalities. This question does not have a generic answer. However, recent studies investigating text processing have demonstrated that low-level motor execution patterns exhibited in participants' typing behaviour can provide a useful window on higher-level processing (Alves et al., 2008; Eklundh & Kollberg, 2003; Sullivan & Lindgren, 2006; Van Waes & Schellens, 2003). Our approach has been to check, wherever possible, people's chat-tool responses against existing spoken language corpora and experimental data. Our results to date indicate that text-based variants of key dialogue tasks, including the tangram task (Clark & Wilkes-Gibbs, 1986), the maze task (Garrod & Doherty, 1994) and the story-telling task (J. Bavelas et al., 1992), exhibit local and global patterns in participants' dialogue that are comparable to those observed in the original spoken versions (Eshghi, 2009; Healey & Mills, 2006; Healey et al., 2003; Howes et al., 2011; Mills & Healey, 2006; Mills, 2013). Although the gross characteristics of different modalities are clearly different, e.g. text persists in a way that speech does not, our experience is that this doesn't substantially alter the use of, e.g., clarification questions, acknowledgments, reformulations, compound contributions or repairs, since these are primarily shaped by the pressure to maintain mutual-understanding.

Conclusion

The 'DiET' platform is a dialogue experimentation toolkit that enables a new approach to fine-grained, word and turn-level interventions in free dialogue. It takes advantage of features of text-chat and natural language processing to provide new ways to investigate the structural, procedural and conceptual organisation of dialogue through the real-time capture, processing and transformation of turns as they are constructed. This approach overcomes some important limitiations of existing experimental techniques. The software runs on Linux, Windows and Apple platforms and is released freely under a General Public License (GPL) at http://cogsci.eecs.qmul.ac.uk/diet/.

References

Alves, R. A., Castro, S. L., & Olive, T. (2008). Execution and pauses in writing narratives: Processing time, cognitive effort and typing skill. *International journal of psychology*, *43*(6), 969–979.

Anderson, A., Bader, M., Bard, E., Boyle, E., Doherty, G. M., Garrod, S., ... Weinert, R. (1991). The hcrc map task corpus. *Language and Speech*, *34*, 351-366.

Arnold, J. E., Fagnano, M., & Tanenhaus, M. K. (2003). Disfluencies signal theee, um, new information. *Journal of Psycholinguistic Research*, *32*, 25-34.

Arnold, J. E., Hudson, K., & Tanenhaus, M. K. (2007). If you say thee uh- you're describing something hard: the on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *5*, 914-930.

Bailenson, J. N., Yee, N., Blascovich, J., & Guadagno, R. (2008). Transformed social interaction in mediated interpersonal communication. In *Mediated interpersonal communication* (p. 77-99). Lawrence Erlbaum Associates.

Bard, E. G., Anderson, A. H., Chen, Y., Nicholson, H. B. M., Havard, C., & Dalzel-Job, S. (2007). Let's you do that: Sharing the cognitive burdens of dialogue. *Journal of Memory and Language*, *57*(4).

Barr, D., & Keysar, B. (2002). Anchoring comprehension in linguistic precedents. *Journal of Memory and Language*, *46*, 937-962.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology* (Vol. 14). Cambridge University Press.

Bavelas, A. (1950). Communication patterns in task-oriented groups. *The Journal of the Acoustical Society of America*, *22*(6), 725–730.

Bavelas, J., Black, A., Lemery, C., & Mullett, J. (1986). I show how you feelŤ: Motor mimicry as a communicative act. *Journal of Personality and Social Psychology*, *50*(2), 322–329.

Bavelas, J., Chovil, N., Lawrie, D., & Wade, L. (1992). Interactive gestures. *Discourse Processes*, *15*, 469-489.

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

Bavelas, J., & Gerwing, J. (2011). The listener as addressee in face-to-face dialogue. *International Journal of Listening*, *25*(3), 178–198.

Bavelas, J., Gerwing, J., Sutton, C., & Prevost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language*, *58*(2), 495–520.

Beattie, G., & Aboudan, R. (1994). Gestures, pauses and speech: An experimental investigation of the effects of changing social context on their precise temporal relationships. *Semiotica*, *99*(3-4), 239–272.

Beebe, B., Alson, D., Jaffe, J., Feldstein, S., & Crown, C. (1988). Vocal congruence in mother-infant play. *Journal of Psycholinguistic Research*, *17*(3), 245–259.

Berko, J. (1958). *The child's learning of english morphology.* Unpublished doctoral dissertation, Radcliffe College.

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive psychology*, *18*(3), 355–387.

Boyle, E. A., Anderson, A., & Newlands, A. (1994). The effects of visibility on dialogue and performance in a cooperative problem solving task. *Language and Speech*, *37*, 1-20.

Bradshaw, B. (2013). Facebook looks to crack chat app riddle -. *Financial Times, April 6, 2013*, 9.

Branigan, H., Pickering, M., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, *75 (B)*, 13-25.

Branigan, H., Pickering, M. J., McLean, J. F., & Cleland, A. A. (2007). Syntactic alignment and participant role in dialogue. *Cognition*, *104*(2), 163–197.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1482–1493.

Brennan, S. E., Galati, A., & Kuhlen, A. (2010). Two minds, one dialog: Coordinating speaking and understanding. *The Psychology of Learning and Motivation*, *53*, 301-344.

Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialog. *Topics in Cognitive Science*, *1*, 274-291.

Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language*, *44*(2), 274–296.

Brown-Schmidt, S., & Tanenhaus, M. K. (2008). Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive science*, *32*(4), 643–684.

Brunyé, T. T., Ditman, T., Mahoney, C. R., Augustyn, J. S., & Taylor, H. A. (2009). When you and i share perspectives pronouns modulate perspective taking during narrative comprehension. *Psychological Science*, *20*(1), 27–32.

Carletta, J., Hill, R. L., Nicol, C., Taylor, T., De Ruiter, J. P., & Bard, E. (2010). Eye-tracking for two-person tasks with manipulation of a virtual world. *Behavior Research Methods*, *42*, 254-265.

Cassar, A. (2007). Coordination and cooperation in local, random and small world networks: Experimental evidence. *Games and Economic Behavior*, *58*(2), 209–230.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social PSychology*, *76*, 893-910.

Chevalley, E., & Bangerter, A. (2010). Suspending and reinstating joint activities with dialogue. *Discourse Processes*, *47*(4), 263–291.

Chouinard, M. M., & Clark, E. V. (2003). Adult reformulations of child errors as negative evidence. *Journal of child language*, *30*(3), 637–670.

Clark, H. H. (1979). Responding to indirect speech acts. *Cognitive psychology*, *11*(4), 430–477.

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

Clark, H. H. (1997). Dogmas of understanding. *Discourse Processes*, *23*, 567-598.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speech. *Cognition*, *84*, 73-111.

Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, *50*, 62-81.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *2*, 1-39.

Doherty-Sneddon, G., Anderson, A., O'Malley, C., Langton, S., Garrod, S., & Bruce, V. (1997). Face-to-face and video-mediated communication: A comparison of dialogue structure and task performance. *Journal of Experimental Psychology: Applied*, *3*(2), 105.

Edlund, J., & Beskow, J. (2009). Mushypeek: A framework for online investigation of audiovisual phenomena. *Language and Speech*, *52*, 351-365.

Eklundh, K. S., & Kollberg, P. (2003). Emerging discourse structure: computer-assisted episode analysis as a window to global revision in university students' writing. *Journal of Pragmatics*, *35*(6), 869–891.

Eshghi, A. (2009). *Uncommon ground: the distribution of dialogue contexts.* Unpublished doctoral dissertation, Queen Mary, University of London.

Eshghi, A., Healey, P. G. T., Purver, M., Howes, C., Gregoromichelaki, E., & Kempson, r. (2010, September). Incremental turn processing in dialogue. In *Architectures and mechanisms for language processing (amlap).* York, UK.

Fay, N., Garrod, S., & Carletta, J. (2000). Group discussion as interactive dialogue or as serial monologue: The influence of group size. *Psychological Science*, *11*(6), 481–486.

Freeman, L. C. (1979). Centrality in social networks conceptual clarification. *Social networks*, *1*(3), 215–239.

Freeman, L. C., Roeder, D., & Mulholland, R. R. (1980). Centrality in social networks: Ii. experimental results. *Social networks*, *2*(2), 119–141.

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive science*, *29*(5), 737–767.

Garcia, A., & Jacobs, J. B. (1999). The eyes of the beholder: Understanding the turn-taking system in quasi-synchronous computer-mediated communication. *Research on Language and Social Interaction*, *32*(4), 337-367.

Garfinkel, H. (1967). *Studies in ethnomethodology.* Englewood Cliffs NJ.

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, *27*, 181-218.

Garrod, S., & Doherty, G. (1994). Conversation, coordination and convention: an empirical investigation of how groups establish linguistic conventions. *Cognition*, *53*, 181–215.

Garton, L., Haythornthwaite, C., & Wellman, B. (2006). Studying online social networks. *Journal of Computer-Mediated Communication*, *3*(1), 0–0.

Ginzburg, J. (2011). *The interactive stance (forthcoming)* (CSLI, Ed.). forthc.

Ginzburg, J., & Cooper, R. (2004). Clarification, ellipsis, and the nature of contextual updates. *Linguistics and Philosophy*, *27*, 297-366.

Goodwin, C. (1979). The interactive construction of a sentence in natural conversation. *Everyday language: Studies in ethnomethodology*, 97–121.

Gratch, J., Okhmatovskaia, A., Lamothe, F., Marsella, S., Morales, M., & Van der Werf, R. J. (2006). Virutal rapport. In *Proceedings of the 6th international conference on intelligent virual agents* (p. 14-27).

Gregoromichelaki, E., Kempson, R., Purver, M., Mills, G. J., Cann, R., Meyer-Viol, W., & Healey, P. G. T. (2011). Incrementality and intention-recognition in utterance processing. *Dialogue and Discourse*, *2*, 199-233.

Hanna, J. E., & Brennan, S. E. (2007). SpeakersŠ eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language*, *57*(4), 596–615.

Healey, P. G. T. (1997). Expertise or expert-ese: The emergence of task-oriented sub-languages. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the 19th annual conference of the cognitive science society* (p. 301-306).

Healey, P. G. T., Eshghi, A., Howes, c., & Purver, M. (2011, July). making a contribution: Processing clarification requests in dialogue. In *Proceedings of the 21st annual meeting of the society for text & discourse.* Poitiers.

Healey, P. G. T., Garrod, S., Fay, N., Lee, J., & Oberlander, J. (2002). Interactional context in graphical communication. In *Proceedings of the 24th annual conference of the cognitive science society* (p. 441-446).

Healey, P. G. T., & Mills, G. J. (2006). Participation, precedence and co-ordination. In *Proceedings of the 28th annual conference of the cognitive science society.* Vancouver, Canada.

Healey, P. G. T., Purver, M., King, J., Ginzburg, J., & Mills, G. J. (2003). Experimenting with clarification in dialogue. In *Proceedings of the 25th annual conference of the cognitive science society* (p. 539-544). Mahwah, N. J>.

Healey, P. G. T., Swoboda, N., Umata, I., & King, J. (2007). Graphical language games: Interactional constraints on representational form. *Cognitive Science*, *31*, 285-309.

Heritage, J. (2002). Oh-prefaced responses to assessments: A method of modifying agreement / disagreement. In C. Ford, b. Fox, & T. S. (Eds.), *The language of turn and sequence* (p. 196-224). Ox.

Heritage, J., & Raymond, G. (2005). Navigating epistemic landscapes: Acquiescence, agency and resistance in responses to polar questions. In J.-P. de Ruiter (Ed.), *Questions.* Cambridge University Press.

Herring, S. C. (1999). Interactional coherence in cmc. *Journal of Computer Mediated Communication*, *4*.

Howes, C., Healey, P. G. T., Purver, M., & Eshghi, A. (2012, August). Finishing each otherŠs... responding to incomplete contributions in dialogue. In *Proceedings of the 34th annual meeting of the cognitive science society.* Sapporo, Japan.

Howes, C., Purver, M., Healey, P. G. T., Mills, G., & Gregoromichelaki. (2011). On incrementality in dialogue: Evidence from compound contributions. *Dialogue and*, *2*, 279-311.

Jackson, M., Anderson, A., McEwan, R., & Mullin, J. (2000). Impact of video frame rate on communicative behaviour in two and four party groups. In *Proceedings of conference on computer supported cooperative work.* New York.

Jefferson, G. (1987). On exposed and embedded correction in conversation. *Talk and social organization*, 86–100.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*(31), 10681–10686.

Klein, D., & Manning, C. D. (2003). Accurate unlexicalized parsing. In *Proceedings of the 41st meeting of the association for computational linguistics.* MIT Press.

Kleiner, M., Brainard, D., & Peli, D. (2007). What's new in psychtoolbox-3. In *Perception 36.* ECVP Abstract supplement.

Krauss, R. M., & Weinheimer, S. (1967). Effects of referent similarity and communication mode on verbal encoding. *Journal of Verbal Learning and Verbal Behavior*, *6*, 359-363.

Kraut, R. E., & Fussell, S. (1996). Social psychological models of interpersonal communication. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (p. 655-701). London: London: Guildford Press.

Kraut, R. E., Fussell, S. R., & Siegel, J. (2003). Visual information as a conversational resource in collaborative physical tasks. *Human Computer Interaction*, *18*, 13-49.

Kuhlen, A. K., & Brennan, S. E. (2013). Language in dialogue: when confederates might be hazardous to your data. *Psychonomic bulletin & review*, *20*(1), 54–72.

Kumar, R., Novak, J., & Tomkins, A. (2010). Structure and evolution of online social networks. In *Link mining: Models, algorithms, and applications* (pp. 337–357). Springer.

Lenhardt, A. (2012). *Teens, smartphones & texting* (Tech. Rep.). http://www.pewinternet.org/Reports/2012/Teens-and-smartphones.aspx: Pew, Internet and American Life Project.

Lerner, G. H. (2002). Turn-sharing: the choral co-production of talk in interaction. In C. Ford, F. B, & T. S. (Eds.), *The language of turn and sequence.* Oxford: Oxford University Press.

Lerner, G. H., & Takagi, T. (1999). On the place of linguistic resources in the organization of talk-in-interaction: A co-investigation of english and japanese grammatical practices. *Journal of Pragmatics*, *31*(1), 49-75.

Levelt, W. J., & Kelter, S. (1982). Surface form and memory in question answering. *Cognitive psychology*, *14*(1), 78–106.

Levinson, S. C. (1983). *Pragmatics.* Cambridge: Cambridge University Press.

Macura, Z., & Ginzburg, J. (2008). Dynamics and adaptiveness of meta-communicative interaction in a foraging environment. *Language in flux: dialogue coordination, language variation, change and evolution*, *1*, 41.

MacWhinney, B., Cohen, J., & Provost, J. (1993). The psyscope experiment-building system. *Spatial Vision*, *11*, 99-101.

Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science*, *31*(1), 133–156.

Marin, A., & Wellman, B. (2011). Social network analysis: An introduction. *The Sage Handbook of Social Network Analysis, London, Sage*, 11–25.

Metzing, C., & Brennan, S. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memo*, *49*, 201-213.

Meyer, D. E., Schvaneveldt, R. W., et al. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of experimental psychology*, *90*(2), 227–234.

Miller, G. A. (1995). Wordnet: A lexical database for english. *Communications of the ACM*, *38*(11), 39-41.

Mills, G. J. (2011, July). The emergence of procedural conventions in dialogue. In *In proceedings of the 33rd annual conference of the cognitive science society.* Boston.

Mills, G. J. (2012a). Clarifying reference and plans in dialogue. In *Dkgl 5.* Freiburg.

Mills, G. J. (2012b, October). Exhuming the procedural common ground: partner-specific effects. In *Ampra pragmatics of the americas.*

Mills, G. J. (2013). Dialogue in joint activity: complementarity, convergence and conventionalization. *New Ideas in Psychology*, *in press*.

Mills, G. J., & Healey, P. G. T. (2006). Clarifying spatial descriptions: Local and global effects on semantic co-ordination. In *Proceedings of the 10th workshop on the semantics and pragmatics of dialogue.*

Munro, R., Bethard, S., Kuperman, V., Tzuyin, L., Melnick, R., Potts, C., . . . Tily, H. (2010). Crowdsourcing and language studies: the new generation of linguistic data. In *Workshop on creating speech and language data with amazon's mechanical turk.* Los Angeles, CA.

Murray, C. T., & Trevarthen, C. (1986). The infant's role in mother–infant communication. *Journal of Child Language*, *13*, 15–29.

O'Conaill, B., Whittaker, S., & Wilbur, S. (1993). Conversations over video conferences: An evaluation of the spoken aspects of video-mediated communication. *Human-computer interaction*, *8*(4), 389–428.

Olson, J. S., Olson, G. M., & Meader, D. K. (1997). Face-to-face group work compared to remote group work with and without video. In *Video mediated communication* (p. 157-172). Hillsdale, NJ: Lawrence Erlbaum Associates.

O'Malley, C., Langton, S., Anderson, A., Doherty-Sneddon, G., & Bruce, V. (1996). Comparison of face-to-face and video-mediated interaction. *Interacting with Computers*, *8*(2), 177–192.

Pentland, A. (2004). Social dynamics: Signals and behavior. In *Proceedings of icdl '04.* San Diego.

Phillips, B. C. (2007). *A comparison of autonomous and collaborative models in computer-mediated communication.* Unpublished doctoral dissertation, University of Victoria.

Pickering, M., & Garrod, S. (2004). The interactive alignment model. *Behavioral and Brain Sciences*, *27*, 169-189.

Pickering, M., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *in press.*

Poesio, M., & Rieser, H. (2010). Completions, coordination, and alignment in dialogue. *Dialogue*, *1*, 1-82.

Purver, M., Healey, P. G. T., King, J., Ginzburg, J., & Mills, G. J. (2003, July). Answering clarification questions. In *Proceedings of the 4th sigdial workshop on discourse and dialogue* (p. 23-33). Sapporo.

Richardson, D. C., Dale, R., & Kirkham, N. Z. (2008). The art of conversation is coordination: Common ground and the coupling of eye movements during dialogue. *Psychological Science*, *18*, 407-413.

Sacks, H. (1995). *Lectures on conversation.* Blackwell Publishing.

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

Saxton, M. (1997). The contrast theory of negative input. *Journal of Child Language, 24*, *24*, 139-161.

Scaife, M., & Bruner, J. S. (1975). The capacity for joint visual attention in the infant. *Nature.*

Schegloff, E. A. (1987). Some sources of misunderstanding in talk-in-interaction. *Linguistics*, *25*, 201-218.

Schegloff, E. A. (1992). Repair after the next turn: The last structurally provided defense of intersubjectivity in conversation. *American Journal of Sociology*, *97*, 1295-1345.

Schegloff, E. A. (1996). Confirming allusions: Toward an empirical account of action. *American Journal of Sociology*, *104*, 161-216.

Schegloff, E. A. (2007). *Sequence organization in interaction: Volume 1: A primer in conversation analysis.* Cambridge University Press,.

Schegloff, E. A. (2010). Some other Şuh(m)Ťs. discourse processes. *Discourse Processes*, *47*, 130-174.

Schlangen, D., & Fernandez, R. (2007, August). speaking through a noisy channel: Experiments on inducing clarification behaviour in human-human dialogue. In *Proceedings of interspeech 2007.* Antwerp, Belgium.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime user's guide.* Pittsburgh: Psychology Software Tools, Inc.

Schober, M. (1993). Spatial perspective-taking in conversation. *Cognition*, *47*, 1-24.

Schober, M., & Clark, H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, *21*, 211-232.

Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less actionŮcandidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience*, *10*(4), 295–302.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., ... others (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, *106*(26), 10587–10592.

Sullivan, K. P., & Lindgren, E. (2006). *Computer keystroke logging and writing: Methods and applications.* Elsevier.

Swinney, D. A. (1979). Lexical access during sentence comprehension:(re) consideration of context effects. *Journal of verbal learning and verbal behavior*, *18*(6), 645–659.

Tanenhaus, M. K., & Brown-Schmidt, S. (2008). Language processing in the natural world. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 1105–1122.

Tannen, D. (2007). Power maneuvers and connection maneuvers in family interaction. *Family Talk: Discourse and identity in four American families*, 27–48.

Thorndyke, P. W. (1977). Cognitive structures in comprehension and memory of narrative discourse. *Cognitive psychology*, *9*(1), 77–110.

Van Baaren, R. B., Holland, R. W., Steenaert, B., & van Knippenberg, A. (2003). Mimicry for money: Behavioral consequences of imitation. *Journal of Experimental Social Psychology*, *39*(4), 393–398.

Van Waes, L., & Schellens, P. J. (2003). Writing profiles: The effect of the writing mode on pausing and revision patterns of experienced writers. *Journal of pragmatics*, *35*(6), 829–853.

Whittaker, S. (2003). Theories and methods in mediated communication. In *The handbook of discourse processes* (pp. 243–286). Erlbaum.

Wilkes-Gibbs, D., & Clark, H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language*, *31*, 183-194.

Table 1

Example Grounding downgrade. Matcher 2 types a turn that contains a grounding cue "OK". Matcher 1 sees a transformed version of this turn which 'downgrades' the grounding cue into a more tentative "Hmm". The Director sees the original turn.

Director's view		Matcher 1's View	Matcher 2's View
D: `Can you see the large square?`		D: `Can you see the large square?`	D: `Can you see the large square?`
M2: `OK, right next to the top?`	**Intervention** $\Longrightarrow$	M2: **Hmm** `right next to the top?`	M2: `OK right next to the top?`
M1: `Yeah it's the topmost`		M1: `Yeah it's the topmost`	M1: `Yeah it's the topmost`

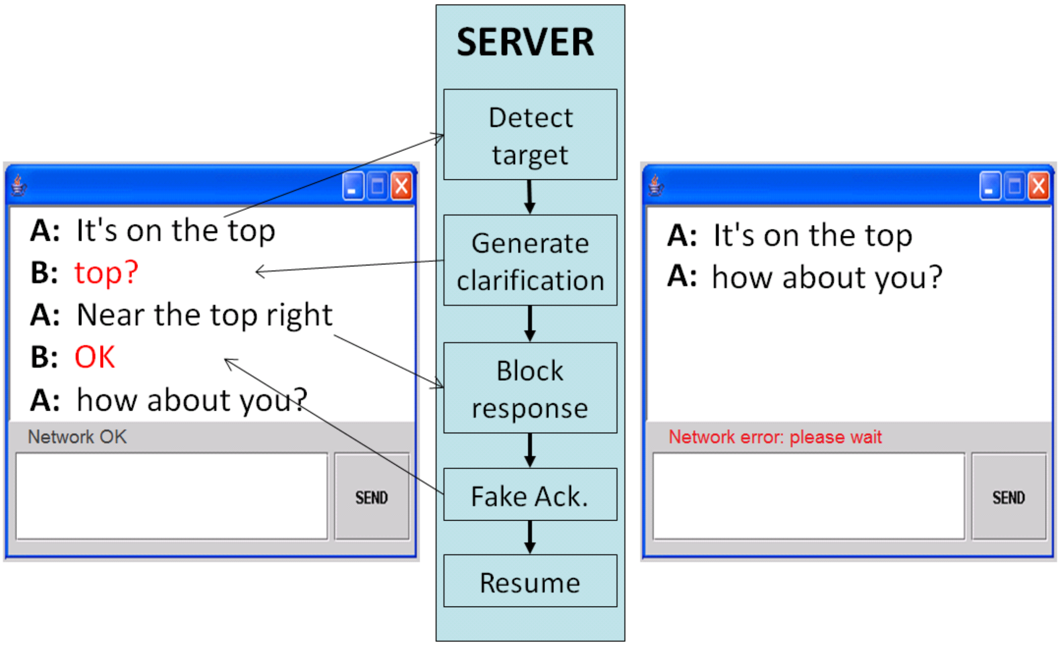

Figure 1. Schematic showing how an intervention unfolds. Two participants are interacting with each other using their chat interfaces. Each interface consists of (1) a main window that displays the dialogue, (2) a status "bar" for displaying information, e.g. whether a participant is currently typing, and (3) a text entry window where participants formulate their contributions. All participants' turns are intercepted by the server. Participant A (left) is responding to an artificially generated clarification request "top?" that A perceives B as having typed. Note here how the clarification is treated by A as a request for elaboration. Importantly, participant B does not see A's response "Near the top right" and also does not see the artificial acknowledgment "OK" that is also generated by the server. Note the artificial error message "Network error: please wait" displayed on B's interface, which gives B a plausible reason for the delay while A engages in the artificial clarification sequence.

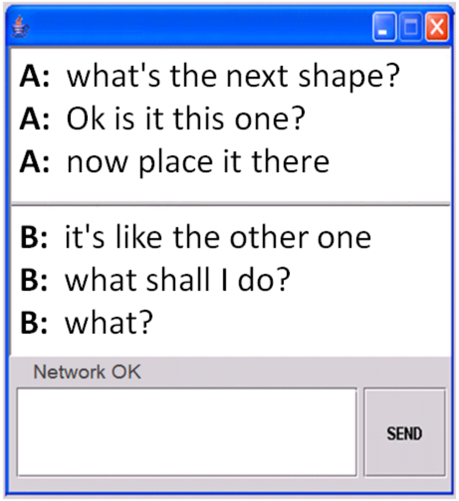

Figure 2. Dual window chat tool. The primary window (1) is sub-divided into separate windows that only display turns from a single participant. The status bar (2) which displays “Network OK" and the text entry window (3) are identical to the single window chat interface shown in Figure 1.

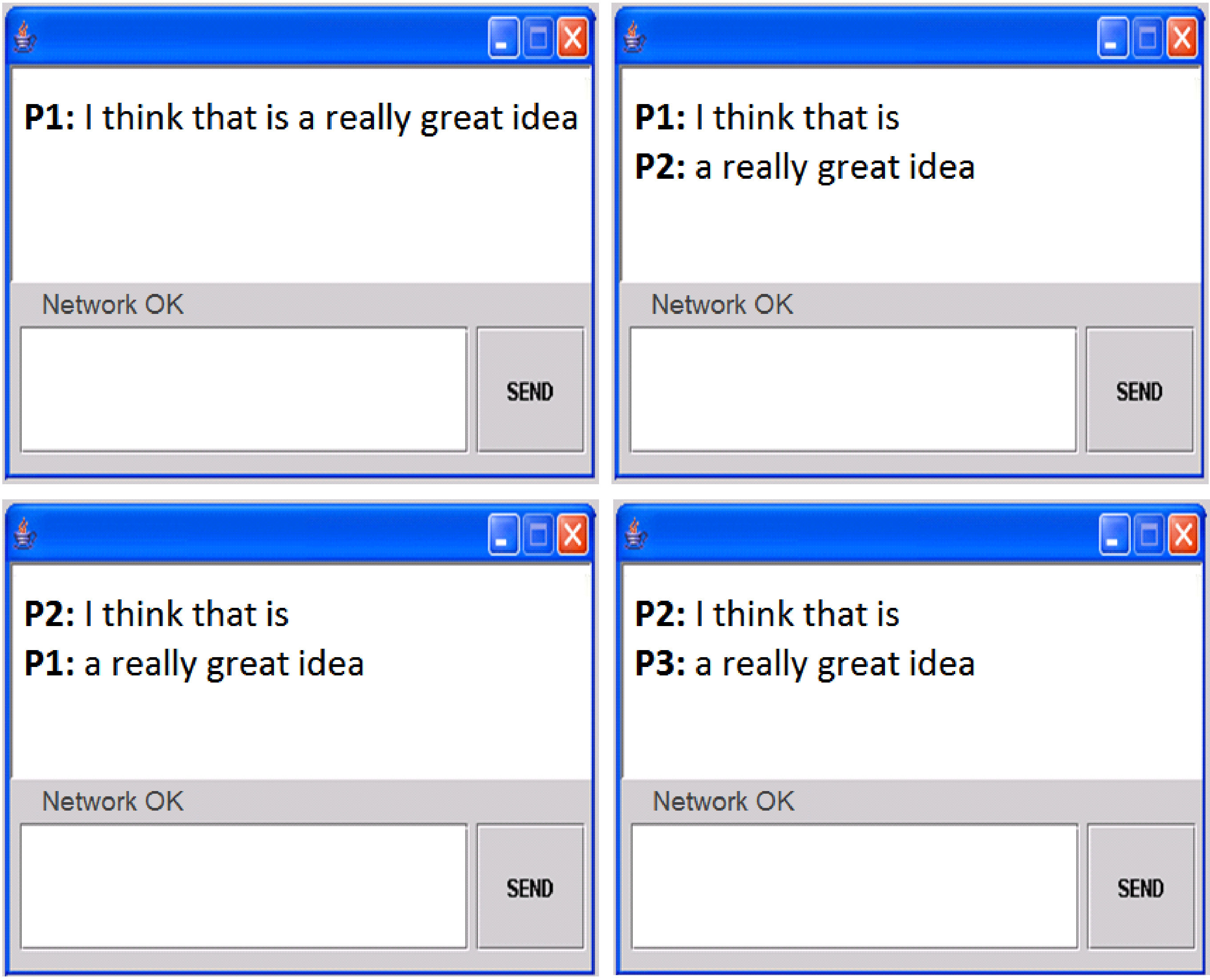

Figure 3. Compound Contributions. The four chat windows correspond to the experimental setup described on page 24. The chat window on the top left is the original text that was typed by participant P1. The chat window on the top right shows *Variant1* which appears to the recipient as if P2 completed P1's turn. The chat window on the bottom left shows *Variant2* which appears to the recipient as if P1 completed P2's turn. The chat window on the bottom right appears to the recipient as if P3 completed P2's turn.