

Language in dialogue: when confederates might be hazardous to your data

Anna K. Kuhlen · Susan E. Brennan

© Psychonomic Society, Inc. 2012

Abstract Experiments that aim to model language processing in spoken dialogue contexts often use confederates as speakers or addressees. However, the decision of whether to use a confederate, and of precisely how to deploy one, is shaped by researchers' explicit theories and implicit assumptions about the nature of dialogue. When can a confederate fulfill the role of conversational partner without changing the nature of the dialogue itself? We survey the benefits and risks of using confederates in studies of language in dialogue contexts, identifying four concerns that appear to guide how confederates are deployed. We then discuss several studies that have addressed these concerns differently—and, in some cases, have found different results. We conclude with recommendations for how to weigh the benefits and risks of using experimental confederates in dialogue studies: Confederates are best used when an experimental hypothesis concerns responses to unusual behaviors or low-frequency linguistic forms and when the experimental task calls for the confederate partner to take the initiative as speaker. Confederates can be especially risky in the addressee role, especially if their non-verbal behavior is uncontrolled and if they know more than is warranted by the experimental task.

Keywords Psycholinguistics · Speech perception · Speech production · Social cognition · Confederates · Dialogue

A. K. Kuhlen (✉)
Berlin School of Mind and Brain,
Humboldt University of Berlin,
10099 Berlin, Germany
e-mail: anna.kuhlen@bccn-berlin.de

S. E. Brennan
Department of Psychology,
Stony Brook University, Stony Brook,
New York 11794, USA
e-mail: susan.brennan@stonybrook.edu

Conversational partners in laboratory experiments are not always what they seem. For example, subjects may think that they are interacting via intercom with another naive subject in the next room when, in fact, they are responding to prerecorded utterances. Or they may think that they are interacting with a computer when, in fact, an experimenter (or a “Wizard of Oz”) is responding to them. Psycholinguistic studies concerned with generalizing to language processing in context are increasingly situated in dialogue, where they are likely to include a conversational partner for the subjects. It is not uncommon for such partners not to be naive, like the subjects, but instead to be accomplices whose purpose it is to assist the experimenter in staging the dialogue context.

The use of confederates is of course a longstanding tradition in social psychology. Often the unusual behavior of one or more people is intended to serve as the main stimulus in the experiment, so staging the situation with confederates may be the only practical way to collect data. Alternatively, using a confederate may simply be a convenience for the experimenter. However, when it comes to serving as conversational partners in language experiments, confederates do not always do a good job. Concerns have been raised as to the validity of studies of communication that have used confederates (e.g., Bavelas, Gerwing, Sutton, & Prevost, 2008; Lockridge & Brennan, 2002; Tanenhaus & Brown-Schmidt, 2007). In experimental examinations of ordinary cognitive processing in social contexts, confederates' behavior (both verbal and nonverbal) may differ systematically from naive subjects' behavior, thereby distorting the interaction and the processes, representations, and behavior under study.

The issue of whether (and how) confederates should replace conversational partners is not merely methodological; it also raises theoretical questions about the nature of dialogue itself. The measures taken to integrate confederates into experiments and to strive for ecological validity can be revealing about how researchers conceptualize communication and about their assumptions regarding the roles of

speakers and addressees in dialogue. Our analysis focuses on dialogue, but it may be relevant as well to experimental protocols in related fields such as collaborative memory (e.g., Ekeocha & Brennan, 2008; Harris, Paterson, & Kemp, 2008; Hollingshead, 1998; Roediger, Meade, & Bergman, 2001; Weldon & Bellinger, 1997), decision making or problem solving in groups (e.g., Kerr & Tindale, 2004), social cognition (e.g., De Jaegher, Di Paolo, & Gallagher, 2010; Smith & Semin, 2004), child development (e.g., Matthews, Lieven, & Tomasello, 2010; Sobel & Corriveau, 2010), and joint motor coordination (e.g., Reed et al., 2006; Sebanz, Bekkering, & Knoblich, 2006; Shockley, Santana, & Fowler, 2003), as well as in some areas of social neuroscience (which Hari & Kujala, 2009, have termed *two-person neuroscience*; see also Cacioppo & Berntson, 2004; Frith & Frith, 2010; Hasson, Ghazanfar, Galantucci, Garrod, & Keysers, 2012; Montague et al., 2002; Schilbach et al., *in press*). The broader question at stake is, *When can studies of cognitive processing in social contexts focus only on one side of an interaction while standardizing the other, without obscuring the phenomena of interest?*

Our goal is to identify the circumstances under which a confederate can effectively fulfill the role of conversational partner in an experiment just as well as a naive subject can.¹ We begin by discussing the reasons, often methodologically motivated, for using confederates in language studies. We then sketch some different (often implicit) theories about the roles played by partners in dialogue. We identify and discuss concerns that have been raised about confederates; some of these concerns are relevant to the use of confederates in psychology research in general, and some are specific to psycholinguistic studies. We illustrate these concerns with several published studies of language processing that have tested similar hypotheses while employing confederates in different ways—and that have found different results. Finally, we suggest some general guidelines for weighing the benefits and risks of deploying confederates in studies of language processing in dialogue.

Why use confederates in language studies?

The decision to use a confederate is linked to the empirical tradition that a researcher follows. Some language researchers shun the use of confederates as conversational partners and collect data only in everyday settings, for example by recording spontaneous conversations on the street (e.g., M. H. Goodwin, 1985), at the dinner table (e.g., C. Goodwin,

1979), or at the auto repair shop (e.g., Streeck, 2003). This approach follows an ethnographic or ethno-methodological tradition (as in the sociolinguistic field of *conversation analysis*) that aims to describe the underlying sequential organization of conversation without imposing control on the data collection or a priori hypotheses on the data (see, e.g., Bergmann, 1981; Deppermann, 2008; Heritage, 1984; and Levinson, 1983, for discussion of conversation-analytical methods). In the conversation analysis tradition, explanations are developed that the participants themselves would probably agree with (Levinson, 1983); in fact, sometimes the researcher is also a participant in the conversation, which would take place whether data were being collected or not (and so, the researcher is not considered to be a confederate). In this tradition, confederates are believed to interfere with the behavior under study: Letting dialogue unfold spontaneously can reveal unexpected patterns of behavior and lead to a better and more accurate understanding of phenomena that might go undetected in more controlled settings.

However, the uncontrolled approach to studying language in dialogue settings comes with a cost: The range of variation in spontaneous language use is enormous, which makes it difficult to compare one conversation to another or to generalize findings beyond a particular set of conversations. In randomly sampled conversations, the communicative intentions of speakers and addressees often can be inferred only retrospectively (and not necessarily accurately) on the basis of their behavior. And, most importantly, collecting dialogue data in an uncontrolled setting does not allow for making predictions or causal inferences about the psycholinguistic and other cognitive mechanisms that underlie observed behavior (where the “behavior” is often reduced to a text transcript of the conversation).

To uncover underlying mechanisms, psycholinguists therefore prefer to study language processing in controlled settings. Conversations in the laboratory are usually more constrained in terms of topic, context, and communicative goals. Typically, subjects are assigned specific experimental tasks (e.g., giving and receiving directions, or describing and identifying objects or abstract shapes) in order to motivate speaking, to limit behavior within a particular domain, and to enable the researcher to time a speaker’s utterances or an addressee’s responses, and perhaps to synchronize them with other behavior. Using task-oriented conversations makes it easier for the observer to detect what interlocutors intend by their utterances (Ito & Speer, 2006; Schober & Brennan, 2003). The challenge of studying dialogue experimentally is to create situations in which language processes can be observed in a controlled fashion while preserving the natural development of the phenomena of interest—essentially, to throw out the bathwater but retain the baby. The bottom line is that control sometimes seems difficult to achieve in a dialogue without the use of confederates.

¹ We use the term *subject* to mean a naive participant from whom data are collected, because confederates (and even the experimenter) may also be construed as “participants” in a dialogue experiment.

Research questions in language studies often focus on *either* the production or the comprehension of language (rather than on dialogue itself), and a researcher so focused may decide that the *other* conversational partner should be a confederate. Keeping one participant's behavior constant and comparable across experimental conditions by having that person be a confederate seems to be a sensible way to establish a certain level of experimental control while still enabling the other (naive) participant to interact and respond spontaneously. The confederate is considered part of the experimental context, ostensibly standardizing the partner's behavior across different interactions (a similar assumption guides standardization in survey research, which aims to avoid bias by presenting each respondent with the same questions read in exactly the same form by an interviewer; for a discussion, see Conrad & Schober, 2000). The same confederate may be used throughout a study, which means that the confederate experiences the same procedures repeatedly. Confederates' behavior may be spontaneous or else scripted, and they may or may not receive detailed instructions on how to behave (apart from "just act natural"). In this way, confederates serve as a sort of stimulus or independent variable, with the naive subject's behavior as the target or dependent variable. In the next section, we will consider how the role a confederate plays in the experiment (if primarily as speaker or addressee) is relevant to trade-offs in how and whether to use a confederate in a language experiment.

Decisions to use confederates in language experiments (as well as in other kinds of experiments) are often driven by the following goals.

Collecting data efficiently A common reason for using confederates is for efficient data collection. Since the availability of research subjects is often limited and dialogue data are labor-intensive to collect, using a confederate instead of a naive conversational partner may be seen as economical: The number of subjects who must be recruited is cut in half. This avoids the considerable time and effort needed to coordinate multiple schedules, as well as situations in which, if one person fails to appear, the experimental session must be cancelled or rescheduled (see Solano, 1989). Using confederates therefore seems to be in researchers' best interests for expediting their studies (and their careers).

Increasing the frequency of rare events In some studies, the behavior of interest occurs in response to a context that is so rare that a very large number of naturally occurring interactions would need to be recorded in order to yield a sufficient number of observations. In social psychological studies of behavior in unusual or nonnormative situations (as in the case of classic studies like those of Asch, 1955; Bandura, Ross, & Ross, 1961; Cozby, 1972; Dutton &

Aron, 1974; Korte, Ypma, & Toppen, 1975; Milgram, 1974; and Page, 1977), this is a compelling rationale indeed; without confederates, there may be no other feasible way to conduct the study. Even conversation analysts have resorted to using confederates when the goal is to study responses to nonnormative behavior; see Garfinkel's "breaching experiments" (Garfinkel, 1952, 1963, 1967). In experimental studies of sentence processing or syntactic ambiguity, psycholinguists may use a confederate in the speaker role so that they can examine the effects of different lexical or syntactic forms on processing (e.g., Branigan, Pickering, & Cleland, 2000; Hartsuiker, Pickering, & Veltkamp, 2004; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Although language studies sometimes place subjects in infelicitous situations, that may not be an aim, but simply the unintended consequence of creating a dialogue context that yields enough responses to test the hypothesis of interest. Sometimes this can be accomplished instead through a judicious choice of the task (see Kraljic & Brennan, 2005, for a discussion).

Reducing exuberant data Many studies of speech *production* depend on eliciting comparable utterances from multiple speakers. Even a simple logical query can be expressed in countless ways. For example, the query *Which programmers work for department managers?* generated 7,000 different formal versions before the researchers stopped counting (see Brennan, 1990). Because spontaneous utterances can be so variable (leading to what Bock, 1996, labeled "exuberant responding"), the challenge becomes how to restrict subjects in what they can say. Experimental investigations that depend on comparing utterances with the same or similar surface structures (e.g., F. Ferreira & Swets, 2005; Haywood, Pickering, & Branigan, 2005; Kraljic & Brennan, 2005) end up having to exclude data that deviate from targeted forms, decreasing the statistical power of the analysis (Bock, 1996). The cost of transcribing hours of naturally occurring conversation in search of targeted forms is considerably greater than the cost of transcribing subjects' responses to a confederate's scripted prompt.

In the face of these needs—to increase the frequency of rare events for comprehension studies and to reduce exuberant data in production studies—the appeal of using confederates is so strong that justification may seem necessary for *not* doing so.

Focusing on the individual as the unit of study Historically, psychology has been defined as the study of the individual (e.g., Allport, 1954, 1969, 1985; but see Vygotsky, 1978. See also Bavelas, 2005, and Solano, 1989, for discussions of how this focus on the individual has affected the field of social psychology). Similarly, the fields of cognitive psychology in general, and of psycholinguistics in particular, tend to study individual minds and, hence, to focus on

language use in isolation (see Sebanz et al., 2006, for a discussion of how this has affected the field of cognitive psychology). When the goal is to understand individual minds or behavior, social context appears to be a confounding variable that needs to be controlled. This rationale also underlies many studies of dialogue.

Going beyond monologue Traditionally, psycholinguists have investigated language processing in monologue settings, whether for reasons of experimental control or because of an implicit assumption that the communicative context of language use (“performance”) doesn’t affect language processing in any interesting way (e.g., Chomsky, 1965, 1980). Perception and comprehension experiments in this tradition have had lone subjects listen to speech that is known to be prerecorded; in production experiments, lone subjects have been made to speak into tape recorders. More recently, there has been a movement to study language processing within dialogue, driven by the prediction that processing may be enhanced by “parity” between comprehension and production when these processes take place in parallel (Pickering & Garrod, 2004). Others have done so for longstanding theoretical reasons, with the assumption that conversation, especially face to face, is the primary setting for language acquisition and use (e.g., Bavelas & Chovil, 2000, 2006; Chafe, 1994; Clark, 1996; C. Goodwin, 1981; Levinson, 1983; Linell, 2005; Tomasello, 2003). Both of these motives lead researchers to include dialogue partners in their experiments; the issue that we address in this study is whether the partner can be a confederate.²

Adhering to the standard statistical tests Studying interactions between individuals can be a challenge for statistical data analysis. Standard statistical analyses, such as analyses of variance and standard regression, are not designed to study dyadic or group processes (Kenny, 1996; Kenny, Kashy, & Cook, 2006). These commonly used analyses rely on the statistical independence of the data. However, the relationship between conversational partners generally does not follow a linear causal sequence in which one turn affects only the next (a false assumption that has been termed *pseudounilaterality*; Duncan, Kanki, Mokros, & Fiske, 1984), but instead, conversational partners reciprocally influence each other. Therefore, the behavior of conversational partners cannot be assumed to be independent (in fact, the point of dialogue studies is often to investigate how the behavior of one partner depends on that of the other). Researchers may be motivated to use confederates (and

thereby to focus the analysis on only one of the conversational partners) in order to avoid violating the assumptions underlying the usual statistical analyses.

More recently, alternative approaches to data analysis have been applied that are better suited to studying conversational partners in interaction. For example, one approach is to use multilevel or mixed effect models to account for behavior nested within a dyad or a conversational unit (see, e.g., Kenny & Kashy, 2010; Forster & Masson, 2008). Another approach, informed by nonlinear physics and dynamical systems theory, can quantify the degree of reoccurring patterns within the behavior of two interacting individuals over time (*cross-recurrence quantification analysis*; see, e.g., M. J. Richardson, Marsh, & Schmidt, 2005; Riley & van Orden, 2005; Shockley et al., 2003).

Reducing complexity In order to investigate a complex phenomenon like dialogue, researchers often strive to reduce it to more basic features or subprocesses (for a detailed critique, see Bavelas, 2005). Taken to the extreme, the reductionist approach suggests that complex behavior such as conversation can be modeled by the sum of its parts (e.g., the contributions of each individual partner). On that assumption, the rationale for using a confederate is to isolate the basic constituents of dialogue by attempting to hold constant the behavior of one conversational partner.

How confederates are deployed reveals implicit theories or assumptions about dialogue

Although many agree on the relevance of studying language processes in dialogue (as opposed to exclusively in monologue), there is disagreement about the nature of dialogue, and therefore about what exactly about dialogue needs to be reproduced in the lab in order for an experiment to achieve ecological validity.³ In particular, researchers differ in what they believe that the role of a conversational partner is; this affects how much thought they give to how to integrate a confederate into this role. For example, some include a conversational partner primarily to give the experimental situation the appearance of being more like a dialogue, whereas others are interested in how a conversational

² Determining whether a particular psycholinguistic experiment should include some form of (real, confederate, or simulated) dialogue partner rather than focusing on autonomous processing depends on what the experiment is trying to model, and this question is beyond the scope of the present article.

³ Not every type of research may strive for ecological validity. For a discussion weighing the costs and benefits of ecological validity and experimental control in the field of memory research, see Banaji and Crowder (1989), and for similar discussion in experimental psychology more generally, see Mook (1983). However, when it comes to studying pragmatics and language use in the context of dialogue, we argue that ecological validity is a crucial factor. Independent of how a researcher values ecological validity, a serious risk to any type of research is overlooking confounding factors, which can lead to potentially artifactual conclusions.

partner could directly influence processing. Either way, how researchers integrate a confederate into an experimental setting reflects their theories, implicit or explicit, on the nature of dialogue. Distinct theoretical stances on the role of a partner in dialogue include the following.

The motivational partner This stance is consistent with social facilitation theory, which proposes that the mere presence of an audience improves individual performance (e.g., Triplett, 1897; Zajonc, 1965). In other words, the role of a live dialogue partner is mainly to get subjects into a dialogue “mode.” At the extreme, the partner may be treated as a prop that authenticates the experimental situation as a dialogue, motivating the subject to treat it as one. For example, V. S. Ferreira and Dell (2000, Exp. 6) tested whether “communicative pressures” created by having an addressee would enable speakers to better detect and avoid ambiguity (the addressees, who were not confederates, did not speak, but did rate the speakers’ clarity).

Underlying this view is the assumption that conversational partners participate actively in a dialogue only when they are speaking, with addressees assumed to be relatively passive. This view conceptualizes conversational partners as either “speaker” or “speaker-in-waiting” (for a discussion, see Bavelas, Coates, & Johnson, 2000). Dialogue is seen as a sequence of alternating monologues in which speakers take turns talking. Underlying this perspective seems to be the classic *message model* (Akmajian, Demers, & Harnish, 1987; Shannon & Weaver, 1949), which conceptualizes dialogue as a unidirectional transmission of information from sender to receiver: While the sender encodes a message and speaks, the receiver listens passively and decodes the message while awaiting a speaking turn. The dialogue context is presumed to be more or less static, and the moment-by-moment planning and articulation of an utterance is presumed to originate from cognitive processes and representations that operate autonomously in speaking and in listening.

For an experimenter who takes this view, using one or more confederates instead of naive conversational partners is sometimes seen as a convenient solution, since a confederate can easily fill the motivational role of dialogue partner. When the goal of the experiment is to study production, a confederate addressee is presumed to turn a monologue into a dialogue by their mere presence; the addressee functions as a projection space for a speaker’s utterances. Likewise, when the goal of the experiment is to study comprehension, a confederate speaker is presumed to standardize the context by following a script, without regard to the naive addressee’s behavior. Experimenters who see dialogue partners as primarily motivational often go to great lengths to lead their subjects to believe that the confederate partner is just another

naive subject like themselves. We will discuss concerns with *the covert confederate* approach presently.

The collaborative partner Another view is that dialogue involves more than taking turns speaking and listening: Rather, the interlocutors are mutually responsible for coordinating meaning. They do this by making contributions that are highly contingent and precisely timed, and that provide evidence about mutual understanding and uptake, in a process known as *grounding*. This and other compatible views emphasize collaboration as the essence of dialogue (e.g., Bavelas & Coates, 1992; Clark, 1996; Clark & Wilkes-Gibbs, 1986; Fussell & Krauss, 1989; Roberts & Bavelas, 1996; Schober & Clark, 1989).

Such collaboration can be coordinated to a rather high degree (e.g., Bangerter, 2004; D. C. Richardson & Dale, 2005; D. C. Richardson, Dale, & Kirkham, 2007) and, in fact, with parallel contributions from both partners. An utterance unfolding spontaneously has the potential to be jointly constructed, as when one partner completes what another begins (Wilkes-Gibbs, 1986), or when a speaker adjusts midutterance to visual feedback from an addressee (Brennan, 2005; Brown-Schmidt & Tanenhaus, 2008; Clark, 1996; Clark & Brennan, 1991; Clark & Krych, 2004). According to this view, because conversational partners shape each other’s behavior in a dynamic and reciprocal fashion, the use of a confederate partner in a dialogue experiment could be problematic and needs to be carefully weighed; if the confederate is prevented from responding or is unable to depart from a script, the language game represented by the experiment may be invalid (as it would not approximate natural communication). Confederate speakers who ignore addressees’ needs for clarification or who behave in inauthentic or unexpected ways may yield different patterns of data than do those who act as more authentic speakers (for discussions, see Kuhlen, 2010; Kuhlen & Brennan, 2010; Schober, Conrad, & Fricker, 2004). Authentic speakers are assumed to take into account their addressees’ informational needs and to tailor their utterances accordingly (e.g., Bell, 1984; Clark & Carlson, 1982; Clark & Murphy, 1983; Fussell & Krauss, 1992; Lockridge & Brennan, 2002).

Addressees, as well, display their needs, comprehension, and uptake to speakers via feedback during the grounding process; this can occur through both verbal and nonverbal behavior. Addressees can also make significant contributions to speakers’ utterances as co-creators or co-narrators (for examples, see Bavelas et al., 2000; Krauss, 1987; Wilkes-Gibbs, 1986). So, when conversational partners are not speaking, they are not assumed to be passive recipients; even when they are “just” listening, they can actively shape the interaction. According to the collaborative-partner view, using a confederate, especially in the addressee role, is a

potential minefield if the confederate behaves inflexibly and inauthentically.

The egocentric partner Contrasting with the collaborative view is the theory that conversational partners adapt to each other only in a secondary process. According to this two-stage theory, language processes are egocentric during initial processing, meaning that they do not take the needs or perspectives of conversational partners into account until later in processing (e.g., Barr & Keysar, 2005; Keysar, Barr, Balin, & Brauner, 2000; Keysar, Barr, Balin, & Paek, 1998; Kronmüller & Barr, 2007; see also V. S. Ferreira, Slevc, & Rogers, 2005, as well as Pickering & Garrod's, 2004, notion of "full common ground"—which mandates late processing—and Bard et al.'s, 2000, "dual-process theory"—which mandates late processing for some kinds of linguistic information). Therefore, speakers initially plan utterances independently from any partner-specific knowledge or cues about their addressees' informational needs, and addressees interpret utterances independently from the speakers' perspectives or communicative intentions (guided only by addressees' own perspectives or intentions). If necessary, utterances or interpretations can be adjusted to a partner's needs or common ground, but this occurs only later, after additional inferences, as a repair, or after replanning or reprocessing.

An implicit assumption of this modular view is that "core" psycholinguistic processes unfold no differently in dialogue than in monologue. From this assumption, it follows that a conversational partner becomes relevant only if a researcher is interested in the secondary processes of interactive repair. In this case, the main role of an addressee would be to provide feedback about current understanding, and the role of a speaker would be to correct an utterance if a conversational partner appears to have misunderstood. If the primary research interest was how language is processed in the individual mind (during either comprehension or production), modeling partner-specific effects would become relatively unimportant, as they would be presumed to happen later. As with the motivational-partner view, the mere presence of a partner would be sufficient to stage a dialogue, and so, for reasons of control and convenience, a confederate is often employed.

The interactively aligned partner A proposal related in some ways to the egocentric view, but that considers dialogue to be fundamentally different from monologue, is that having a conversational partner changes the core processes underlying language production and comprehension. This proposal assumes that a tight coupling ("parity") exists between speaking and listening when an individual must be ready to both produce and interpret language in the same context (presumably in contrast to situations such as silently

reading text or speaking aloud in a nondialogic psycholinguistic experiment). According to the interactive-alignment proposal (Pickering & Garrod, 2004), a speaker's utterances automatically activate the same ideas, words, or syntactic structures in the mind of an addressee through a fast, inflexible (and therefore "dumb") priming process. For example, if one partner produces a certain syntactic structure, the other is likely to produce the same structure simply because it has been primed; both partners should subsequently process this structure with more fluency. Dialogue partners' mental representations therefore end up aligning automatically (as opposed to through an active process of jointly constructing meaning during the grounding process, as is proposed by the collaborative view). Although the interactive-alignment view strongly advocates studying linguistic processing in dialogic rather than monologic settings, it deemphasizes the social nature of dialogue and the moment-by-moment coordination among partners. That is, if coordination is the product of an automatic cognitive mechanism, conversational partners have relatively little to do: Speakers merely prime their partner's behavior, and addressees need only listen passively. So, as in the motivational stance, according to the interactive-alignment view a partner should be present, but a confederate could readily take over this role, as the partner's behavior is less important than the speaker's. Experimenters who take this view may employ a confederate for efficiency or, if the hypothesis predicts priming between the subject and partner, as the sort of controlled stimulus described earlier.

Concerns, risks, and findings

As we noted earlier, many classic studies in social psychology have relied on the use of experimental confederates, whether to stage unusual social situations, to induce moods, or to examine the influence of group dynamics on individual behavior, sometimes deceiving and manipulating naive participants by giving false information or intentionally deviating from behavioral norms (e.g., Asch, 1955; Cozby, 1972). Legitimate concerns about this use of confederates have centered around subjects who may become suspicious of a confederate's behavior (see, e.g., Bruehl & Solar, 1970; Martin, 1970, 1973; Orne, 1962; Stricker, Messick, & Jackson, 1967, 1969); often the data must be discarded if there is evidence that a subject does not believe the experiment's cover story (e.g., in a replication of Asch's classic conformity study, 39 %–61 % of the subjects correctly guessed the purpose of the experimental procedure; Stricker et al., 1967). Due to their focus on odd or infelicitous social situations, these classic studies would have been impossible (or at least very difficult) to conduct without

confederates, and the deception about the confederates' status was crucial.

In contrast, many recent studies of language use and language processing have employed confederates to simulate ordinary, natural situations rather than exotic or (intentionally) infelicitous ones. We argue here that the trade-offs and concerns that are relevant to language studies (and to studies of other kinds of cognition in ordinary social contexts, such as collaborative memory in groups) differ from those underlying traditional social psychology studies of nonnormative behavior, and that confederates may sometimes (but not always) add more risk than control to the experimental design. In this section, we first consider two concerns about confederates that are general in nature, and then turn to two concerns that are more specific to studies of dialogue, especially when the scientific questions concern spontaneous adaptation between interacting partners. To illustrate these concerns, we describe in detail several language experiments that have addressed these concerns in different ways and discuss potential influence of such decisions for their results and findings.

Concern 1: the biased confederate One classic concern with using confederates has been raised by work on experimenter bias (Friedman, 1967; Rosenthal, 1966). According to this work, confederates' own expectations about the outcome of a study may cause them to inadvertently bias participants in favor of the experimental hypothesis. The power of confederates to bias behavior is a particular danger when the confederates know what type of behavior is expected from the subjects. It is even more problematic when the confederates know which experimental condition they are participating in or which of their behaviors are predicted to shape the critical or baseline trials. Even the most conscientious confederates are at risk of inadvertently shaping participants' behavior by giving verbal backchannels or nonverbal cues such as facial expressions, body posture, tone of voice, pauses, or eye gaze patterns. Research has shown that these types of cues from an experimenter or examiner can influence people to do better on IQ tests (Congdon & Schober, 2002), inspire children to excel in their schoolwork (Rosenthal & Jacobson, 1968, 1992), lead infants to discard principles of object permanence (Topál, Gergely, Miklósi, Erdőhegyi, & Csibra, 2008), and even cause horses to behave as if they can read and do math (Pfungst, 1907).

To avoid shaping the results, confederates should therefore have as little information as possible about the purpose and hypotheses of the experiment. Ideally, confederates should be blind to the study design, what condition(s) they are participating in, and if possible, the ways in which their behavior may relate to the variables. Sometimes, however, the experimental procedure presents obstacles to this ideal. In those cases, researchers may try to prevent confederates

from leaking information by regulating their nonverbal and verbal behavior. For example, confederates' head movements or eye gaze can be scripted or occluded so as not to cue participants' responses (e.g., Barr & Keysar, 2002; Hanna & Brennan, 2007; Metzging & Brennan, 2003), or confederates may be trained to use the same intonation contour for their utterances across experimental conditions (e.g., Haywood et al., 2005). The confederates' verbal behaviors can be scripted or prerecorded in an attempt to prevent them from treating the subjects differently in different experimental conditions (e.g., Barr & Keysar, 2002; Kronmüller & Barr, 2007), although scripting raises additional concerns that we will discuss presently (see Concern 4 below).

Consider a psycholinguistic study by Keysar et al. (1998, Exp. 2), in which they investigated whether addressees interpret referring expressions using only information that is shared with the speaker (as opposed to information known only to the addressee). In this study, the partner (a confederate who was assumed by the subject to be a fellow naive subject) asked the subject questions about a picture (e.g., a picture of an airplane) that was visually present to both, but that lacked details in the confederate's picture that the confederate needed. In the critical trials, the confederate used an ambiguous referring expression. Just previous to this question, distracter instructions played through the subjects' earphones had directed their attention toward a different, "privileged" object that was occluded from the confederate partner (e.g., a picture of a bird). The confederate's subsequent referring expression was scripted to be either unambiguous or ambiguous, such that it could potentially refer to either the shared or the privileged object (e.g., "Its wings, what color are they?"). Eye gaze was recorded to determine whether the subjects would consider the privileged object (bird) as being a possible referent or else would restrict interpretation to the object that was visually shared with the partner (airplane). The subjects took longer to gaze at the shared referent when there was competition from privileged referents, leading the authors to conclude that reference resolution is not restricted to mutual knowledge, but initially is egocentric.

The confederate in this study was not blind to the experimental conditions. Unknown to the subject, the confederate could hear the subject's "privileged" instruction in order to be able to precisely time the delivery of her (supposedly unrelated) critical instruction; he or she also knew whether or not that instruction would be ambiguous to the subject. The confederate therefore could have behaved somewhat differently across conditions (e.g., perhaps giving subjects less time to recover from the distracter in the ambiguous than in the nonambiguous condition; listeners have, after all, been shown to be quite sensitive to the latency before a speaker's utterance; Brennan & Williams, 1995; Swerts & Krahmer, 2005). Keysar et al. (1998) addressed this

possibility post hoc by comparing the latencies from the end of the distractor instruction to the onset of the critical instruction for the experimental (ambiguous) and baseline (unambiguous) conditions. Finding no significant differences, the authors rejected the possibility that the confederate had biased the results. While it is valuable to rule out biases of known features such as latency, many more features, both unknown and known (e.g., tone of voice, intonation, stress pattern, intelligibility), could remain uncontrolled. An alternative approach (see Concern 3) would be to ensure that confederates' knowledge is appropriate to their role in the task, so that these features would unfold authentically rather than requiring the confederates to be good actors.

Concern 2: the covert confederate Awareness may lead to bias, not only on the part of the confederate who may know too much about the experimental hypotheses or conditions, but also on the part of subjects who may act differently toward known or suspected confederates than toward other naive subjects like themselves. A common concern is that subjects who are aware of the true role of the confederate in the experiment might experience experimental demand that influences them to behave in accord with the assumed hypothesis rather than spontaneously (Bruehl & Solar, 1970; Orme, 1959, 1962, 2002). This issue seems to have concerned dialogue researchers at least as often as the previous issue (the biased confederate), and so some (like their social psychology colleagues before them) have gone to great lengths to hide the true role of the confederate from subjects.

In some studies of social behavior, particularly when the topic concerns responses to the nonnormative behavior of others, the success of an experimental manipulation achieved through a confederate's unusual behavior may well be in question if the naive participant is aware of the confederate's role as accomplice of the experimenter (Martin, 1970, 1973). In fact, research participants who become skeptical about the confederate's true role can behave quite differently from those who do not (Stricker et al., 1969). A confederate who is not credible therefore poses a potential threat to the validity of such a study. Another aspect of this concern is that even if subjects are not deceived about a confederate's status, knowing that they are interacting with a confederate might cause them to become apprehensive about being evaluated (Rosenberg, 1965).

Experimenters often go to great lengths to conceal the status of a confederate. Often, confederates are recruited from a population similar in age to the participants and are trained to pretend to be regular, naive subjects in several ways. Before the actual experiment, elaborate preexperimental encounters may be staged to ensure that the naive subjects assume that the confederate is also a naive participant. Confederates may deliberately arrive late to the experimental session (e.g., Barr & Keysar, 2002), or the

experimenter may display an overt effort to learn the confederate's name (e.g., Branigan, Pickering, McLean, & Cleland, 2007). Although the research question usually predetermines the roles that the confederate and subject play in the experimental task (e.g., as director/instruction-giver or matcher/instruction-follower), the experimenter may pretend to assign these roles in a random manner (e.g., Keysar et al., 1998, 2000). To pretend to be naive about the experimental task, the confederate may ask for clarification of the instructions (e.g., Branigan et al., 2007) or display signs of uncertainty during the experimental session, deliberately interjecting utterances with hesitations, and even making occasional errors (e.g., Branigan et al., 2007; Keysar et al., 2000). At the end of the experimental session, the credibility of the confederate is usually checked via postexperimental questionnaires (e.g., Hanna, Tanenhaus, & Trueswell, 2003; Haywood et al., 2005; Keysar et al., 1998; Keysar, Lin, & Barr, 2003). Sometimes experimenters even offer subjects financial incentives after the study if they are able to correctly guess whether their partner was a confederate or a naive participant (e.g., Keysar et al., 1998, 2000).

Covert confederates vary in how successful they are at deceiving the subjects. Some studies have reported that none of the subjects guessed the confederate's status (e.g., Haywood et al., 2005; Keysar et al., 1998, 2003); others have reported that only a few were suspicious. But, given that these numbers are based on self-report, which is notoriously malleable by various factors (e.g., subjects' diligence, the phrasing of the question, and the assumed social desirability of the answer), the credibility of confederates is likely to be estimated inaccurately. Subjects who do report having been suspicious of the confederate are typically excluded from further analysis (e.g., Keysar et al., 1998; Roediger et al., 2001). Occasionally a large number of subjects correctly guess the confederate's status; in one study, 47 % reported that their conversational partner was a confederate after being offered a financial incentive for correctly guessing the confederate's status (Barr & Keysar, 2002, Exp. 2). The authors compared data from the subjects who had correctly guessed the confederate's identity with those from subjects who had not guessed the confederate's identity. Since no pattern emerged, the authors combined both groups of participants and proceeded with the analysis of the entire data set. This suggests that going to great lengths to conceal the confederate's status may be unnecessary, at least in studies of normative communication.

In fact, some researchers make no attempt to conceal the confederate's status,⁴ under the assumption that someone

⁴ In fact, researchers tend not to conceal confederate partners' status in experiments with child subjects; in first-language acquisition experiments, the experimenters themselves, or even puppets, often act as the (essentially, confederate) partners.

affiliated with the lab can be an authentic interacting partner. A study by Hanna et al. (2003, Exp. 1) tested hypotheses similar to those studied by Keysar et al. (1998, Exp. 2), about the extent to which addressees consider only information that is shared with the speaker when interpreting referring expressions. In Hanna et al.'s experiment, each subject was introduced to a confederate who was accurately identified as a lab assistant. The task of the confederate was to instruct subjects on how to arrange a set of shapes on a display. To standardize the confederate's instructions, specifically with respect to the forms of referring expressions, the confederate used a preformulated script and was trained to sound as natural as possible. In the target trials, the confederate gave ambiguous references that could potentially refer to two different shapes, of which either both were visually copresent to the confederate and the subject, or only one was visually copresent to both participants. Potential interference in reference resolution was measured by tracking the subjects' eye gazes to the objects not visible to the confederates.

In contrast to Keysar et al.'s (1998) Experiment 2 (in which the confederate surreptitiously heard information through headphones that was supposedly privileged to the subject), the privileged object in Hanna et al. (2003) was indeed truly hidden, and therefore unknown, to the confederate. Hence, with the exception of the trials in which the objects were visually present to both, this confederate did not know whether or not the instructions were ambiguous to the subject. At the end of the experiment, the credibility of the confederate's limited knowledge was confirmed with the subjects, who reported that although they thought that the confederate was experienced with the task, they did not think that the confederate knew about the objects that were supposed to be (and actually *were*) visible only to the subjects themselves. The results showed that in the critical trials, subjects were more likely to look at an object in the common ground than at one in privileged ground. Hanna et al. concluded that common ground guides reference resolution from the early moments of processing, and that referent resolution is not egocentric, but instead takes the partner's perspective into account.

The difference in the outcomes of these two studies is striking: Keysar et al. (1998, Exp. 2) concluded that addressees, at least initially, interpret utterances from a privileged, egocentric perspective rather than from their partners' perspectives, whereas Hanna et al. (2003) concluded that addressees can distinguish early on whether information known to them is also known to their partners, and that they take their partners' knowledge into account. Since the two studies differed on several methodological dimensions, a direct comparison is difficult. However, these studies differed in *how* they used confederates: Keysar et al. (1998) carefully kept the status of the confederate hidden, while

Hanna et al. openly informed subjects of the confederate's affiliation with the laboratory. If the status of the confederate were the key influence here, the results from these studies would be expected to have gone in the opposite directions. That is, Hanna et al.'s subjects might have attributed *greater* knowledge to their overt confederate partners, freeing them to ignore their partners' knowledge needs, whereas Keysar et al.'s (1998) subjects (to the extent that they were successfully deceived that the confederate was another subject like themselves) might have been less egocentric. That these results in fact went in the opposite directions suggests that the covertness of a confederate was not the issue.

Perhaps more importantly, these studies also differed in what the confederates *knew*: Keysar et al.'s (1998) confederates were aware of what the subjects could see (although the subjects were deceived about this), while Hanna et al.'s (2003) confederates never had more knowledge than the subjects assumed that they did (this was confirmed by a postexperimental questionnaire). This issue may in fact be what underlies the difference in findings, which brings us to the next concern: Confederates often have more knowledge about the task context than the naive conversational partner expects them to.

Concern 3: the know-it-all confederate Conversational partners should, in theory, avoid informing each other of things that they already know, unless they mark this as shared, given, or definite information. They assess each other's informational needs and adjust their utterances (and other aspects of their communicative behavior, such as gestures) accordingly (e.g., Galati & Brennan, 2006, 2010; Holler & Wilkin, 2009). A conversational partner's presumed knowledge is thereby often linked to his or her role in the task. Speakers, when informing or instructing, are supposed to know more than addressees; when asking a question, they are supposed to know less (so, in terms of this concern, Keysar et al.'s, 1998, confederates knew too much). However, confederates often have an informational advantage inconsistent with their task role. This can result either from their potentially biasing insights into the experimental procedures (as discussed earlier in Concern 1) or simply from repeated experience with the experimental task that leads them to already know what the task requires the subjects to tell them. When confederates' knowledge does not match the knowledge consistent with their role in a conversational task, they may elicit unexpected behavior from their conversational partners (or else expected behavior, but for the wrong reasons).

We note that this can be especially problematic when the confederates are addressees. Consider a study by Brown and Dell (1987) in which they investigated whether speakers design utterances with addressees' informational needs in mind. Their subjects told confederate addressees a series of

short stories in which a main character used either a typical or an atypical instrument to perform a target action (e.g., using a knife or an ice pick to stab someone). During the retelling, some of the addressees followed the narration with the help of illustrations, while others had no illustrations. Hence, the addressees were either aware or unaware of the instrument that the main character in the story had used. Brown and Dell were interested in whether speakers would mention the atypical instruments less explicitly when they knew that their addressees had visual evidence about the instruments. The researchers found that the speakers mentioned atypical instruments more often and earlier in the sentence than they mentioned typical instruments (the latter could to some extent be inferred from the verb). However, whether or not the addressees had illustrations of the instrument did not affect the rate at which speakers mentioned the atypical instruments. Brown and Dell concluded that speakers design their utterances depending on what is easiest for themselves (which also happens to be helpful for addressees in a generic context), and that they consider an addressee's specific informational needs (e.g., knowing or not knowing what instrument was used) *only* in a secondary process, as a repair or afterthought.

However, the (two) confederate addressees in this study had much more knowledge about the stories than their conversational role would have justified. In fact, they heard the stories over and over again throughout the (80) experimental sessions, and most certainly knew the stories better than the subject-storytellers themselves. Throughout the course of the interaction, the confederate addressees might have inadvertently conveyed this to the subject speakers through any feedback that they gave during the interaction (the study did not attempt to specify, or to measure post hoc, the confederates' feedback behavior, although it appeared to the experimenter to be natural; P. M. Brown, personal communication, 1999). If feedback has implicit effects on perception, it is certainly possible that the speakers might not have adapted to their addressees' informational needs, because the addressees did not have any informational needs.

This possibility was raised in a study by Lockridge and Brennan (2002), who replicated Brown and Dell's (1987) study with naive rather than confederate addressees. The speakers in this study showed a different behavioral pattern: With naive addressees who did not have access to additional information through the illustrations, the speakers were more likely to mention atypical instruments, mentioned them earlier in the sentence, and tended to mark them as indefinite. When these addressees had access to the illustration, the speakers were less likely to mention atypical instruments, mentioned them later in the sentence, and tended to mark them as definite. In contrast to Brown and Dell, Lockridge and Brennan therefore concluded that speakers do adjust to their addressees' needs early in utterance planning when their addressees have actual needs.

The contrast in the findings between these two studies is particularly relevant, since the experimental protocols were virtually identical, with the exception that in one the conversational partner was a confederate, and in the other a naive subject. Taken together, these studies suggest that speakers are sensitive to their addressees' behavior, and subsequently adjust their own behavior; when the addressees are perceived as not having any informational needs (e.g., as with confederates who are very experienced with the task), speakers attenuate the information that they provide. This interpretation is consistent with studies by Bavelas et al. (2000) and Kuhlen and Brennan (2010), who showed that speakers narrate stories less vividly when talking to a distracted addressee (who is actually doing a secondary task) than when talking to an attentive addressee. If confederates have unwarranted experience with the conversational task (as they did in Brown & Dell's, 1987 study) or are provided with unwarranted insight into the task (as in Keysar et al.'s, 1998 study, discussed previously), they may appear more knowledgeable to their conversational partners than they would be expected to be, given their role in the conversation. The use of confederates under these circumstances can therefore distort the conclusions.

Concern 4: the scripted confederate In some studies, the utterances of confederate speakers are directed using preformulated scripts. This practice seeks to address concerns such as confederate bias or the variability in utterances and behavior; when scripting is done in sufficient detail, it might prevent knowledgeable confederates from leaking information about the experiment's hypothesis or the experimental task. For studies of speech production, scripting what the confederate says (e.g., Branigan et al., 2000) is a convenient way to try to prompt naive speakers to spontaneously produce utterances in a targeted form, as opposed to "exuberantly." For studies of speech comprehension, standardizing stimulus utterances is especially important when the subjects' reactions are being recorded with time-sensitive measures, such as eyetracking or electroencephalography (EEG), relative to a particular point in the utterance (e.g., the point at which an anomaly arises or ambiguous utterances become disambiguated; see Dahan, Tanenhaus, & Chambers, 2002; Tanenhaus et al., 1995; van Berkum, 2012).

However, using scripted rather than spontaneous utterances raises additional concerns. Preformulated utterances may not sound as natural as utterances occurring in spontaneous dialogue in terms of word choice, syntax, prosody, or articulation—especially if confederates read scripts aloud rather than speaking spontaneously or from memory. Scripted utterances may sound wooden or nonspontaneous, may imply unintended meanings, or may violate pragmatic principles (Bless, Strack, & Schwarz, 1993; Brown-Schmidt & Tanenhaus, 2008). Researchers have adopted different

strategies to address these concerns, such as basing scripts on utterances found in natural conversations (e.g., Brown-Schmidt, 2012; Brown-Schmidt, Gunlogson, & Tanenhaus, 2008, Exp. 2; Metzging & Brennan, 2003). Utterances can be scripted to contain deliberate errors, hesitations, or speech repairs modeled on those produced by naive participants in previous experiments doing similar dialogue tasks (e.g., Branigan et al., 2007). This method may make it less likely that the expressions used by confederates will be perceived as pragmatically infelicitous.

Sometimes spontaneous utterances (with naturally occurring disfluencies) are prerecorded and used as stimuli, sometimes in edited form (e.g., Arnold, Tanenhaus, Altmann, & Fagnano, 2004; Brennan & Williams, 1995; De Ruiter, Mitterer, & Enfield, 2006; Fox Tree, 1995); this approach may suffice to preserve naturalness if the phenomenon of interest can be assumed to be processed autonomously in the same way in which it is processed in interactive dialogue. But, as we will discuss below, prerecorded utterances may well lead to different results than those produced spontaneously (Brown-Schmidt, 2009; see also Schober & Clark, 1989, and Wilkes-Gibbs & Clark, 1992, for evidence that overhearing differs from being addressed).

A second, and perhaps most fundamental, problem with scripted utterances is that they may not be authentically embedded in the ongoing discourse. In spontaneous dialogue, speakers can adapt their utterances contingent on their partners' moment-by-moment behavior, for example by repeating or rephrasing parts of an utterance, or by using the same words or phrases as the partner (Brennan & Clark, 1996; Schober & Clark, 1989). Preformulated utterances lack this flexibility and are not as likely to be contingently aligned with the partner's utterances as spontaneously planned utterances would be. An unintended consequence is that scripted utterances may be understood less well. For example, in survey interviews, questions are understood more accurately when interviewers can spontaneously deviate from a standardized script to provide clarification when they feel their respondents need it (Schober et al., 2004). This suggests that, in order to achieve equivalent levels of understanding across experimental participants, utterances should not be entirely standardized by scripts, but rather should be more flexible. An approach to standardization that is related to scripting has confederates follow rules triggered by the subject's immediately preceding utterance (as in "Wizard of Oz" studies of dialogues with remotely located human or computer partners; see Brennan, 1991; Stent, Huffman, & Brennan, 2008); however, the naturalness of utterances embedded in this way is only as good as the rules themselves.

Interaction between partners is a core aspect of dialogue. A confederate whose behavior is largely scripted (or one whose presence is simulated by prerecorded utterances)

undermines this characteristic. Several dialogue studies have balanced the needs for flexibly contingent interaction, naturalness, and standardization by scripting confederates' utterances only at critical points in the dialogue. Such partial scripting allows confederates to be spontaneous on noncritical trials (e.g., Brown-Schmidt, 2009; Hanna & Tanenhaus, 2004; Metzging & Brennan, 2003), to improvise freely (e.g., Keysar et al., 2003), and to respond spontaneously to requests for clarification from subjects (Conrad & Schober, 2000; Metzging & Brennan, 2003). In Metzging and Brennan's study, for instance, the subjects were told that they were to follow instructions from different directors, and two confederates were employed. First, one of the confederates directed the subjects to identify and arrange a set of objects over three trials (enabling them to entrain on referring expressions for the objects). The confederate and subject were allowed to interact freely, with the exception of the target referring expression, which was scripted. Just before a fourth trial for a set of objects, the confederate paused, said that it might be time for a partner switch, left the room, and either returned, or else the second confederate entered and continued with the task. Critically, in this fourth trial, the (new or old) confederate used either the entrained-upon expression or else a new expression for the target object. The results showed that subjects were slower to resolve new expressions than entrained-upon ones, but only when the new expression was (inexplicably) used by the old partner; a new expression used by a new partner was just as fast to resolve as the old expression (spoken by either partner). This suggests that addressees keep track of the perspectives that they ground with particular partners and interpret referring expressions in partner-specific ways. During this experiment, the confederate was prompted by a booklet that also included the target arrangement for all of the objects, so there was a natural attribution for why the confederate occasionally consulted the booklet (as there was in Hanna & Tanenhaus's, 2004, collaborative cooking task for why the confederate occasionally read from a recipe card).

There is direct evidence that subjects can be sensitive to a partner's spontaneity in dialogue. Two experiments by Brown-Schmidt (2009) directly tested the contribution of spontaneous interaction to speaker-specific processing. The first found a similar result to that of Metzging and Brennan (2003) using a comparable task. In a second experiment, Brown-Schmidt used identical procedures, but in a noninteractive setting, playing aloud to subjects the confederates' utterances that had been recorded in the first experiment. Contrary to the results of the first experiment, the subjects in the second experiment did not differentiate the speaker with whom a referring expression had been entrained: They reacted faster to familiar than to new expressions, regardless of speaker. The difference between these otherwise identical studies was attributed to the lack of interactivity that

resulted not necessarily from using confederates per se, but from using prerecorded utterances (Brown-Schmidt, 2009).

The degree to which confederates' utterances can be scripted or standardized is limited by the modality of the interaction. With prerecorded utterances, there is no physically copresent conversational partner. Scripted confederates may be able to have visual copresence with their subject-partner if they can conceal that they are reading from a script or if there is a good, task-related reason for them to do so (e.g., Branigan et al., 2000; Hanna & Tanenhaus, 2004; Hanna et al., 2003; Metzging & Brennan, 2003; Schoonbaert, Hartsuiker, & Pickering, 2007). Likewise, the modality of the interaction also changes the degree to which confederates' behavior *needs* to be scripted. If the conversation takes place face to face, other observable behaviors of the confederates (e.g., gaze pattern, gestures, and head movements) need to be scripted, predefined, or occluded in a way that seems natural to the situation (e.g., Barr & Keysar, 2002; Metzging & Brennan, 2003). If a conversation is carried out without direct visual contact (e.g., Hanna et al., 2003), as with a remotely located partner who produces text utterances (Brennan, 1991; Healey, Purver, King, Ginzburg, & Mills, 2003), confederates' utterances can be controlled a lot more easily. In addition, if a conversation takes place with text rather than with speech, confederates' utterances need to be less specified in terms of temporal alignment with their partners.

Although researchers are likely to give careful thought to designing the behavior of confederate speakers, they are more likely to neglect the behavior of confederate addressees, rarely scripting or predefineding it in any way (although sometimes examining it post hoc). Again, this neglect probably comes from the widespread, theoretically based assumption that addressees are passive recipients of messages. When confederate addressees' utterances are scripted, the scripts tend to specify only very simple feedback, such as responding "yes" or "no" to the subjects' instructions (e.g., Horton & Keysar, 1996). Scripting addressees' behavior in convincing detail is especially difficult, because this behavior is often largely non-spoken and is contingent to a large degree on what the speakers say and do.

Recommendations

Should researchers avoid using confederates in dialogue experiments? The answer is not a simple *yes* or *no*. How to make the necessary trade-offs about whether and how to deploy a confederate depends to some extent on the research question. On the basis of the concerns raised in the previous section, we offer the following recommendations.

First, to avoid bias and experimental demand, *confederates should be as naive as is feasible to the hypotheses of the*

experiment and the condition that they are currently participating in.

Second, in studies of dialogue, particularly those concerned with questions of whether and how language processing is adapted to the conversational partner, the credibility of confederates may be bolstered *not* by staging additional deception to keep their status hidden, but by *ensuring that subjects attribute the right kind of knowledge to them*. In fact, deception can be counterproductive to this goal, as people are, ironically, *more* likely to leak information when they are explicitly instructed to conceal it (Wardlow Lane, Groisman, & Ferreira, 2006). When a confederate knows too much, naive partners may be able to detect this explicitly or implicitly, and so to adapt in undesirable ways. Thus, in studies of dialogue (unlike of nonnormative situations in social psychology), it is less important that confederates superficially resemble partners who are naive subjects, and more important that their knowledge align with the expectations associated with their role in the interactive task.

A third recommendation concerns how to standardize utterances in dialogue when necessary. When the goal is to study very precise or unusual behavior (e.g., the production or comprehension of specific expressions or rare syntactic forms), the use of confederates may be the only feasible approach. Prerecorded utterances (especially ones that have been produced spontaneously, as opposed to read aloud) may safely serve as the stimuli when the goal is to test autonomous processes of syntactic parsing or certain aspects of comprehension (and may afford more control than scripted utterances performed live, especially if they can be edited). However, embedding prerecorded utterances into a live dialogue when the goal is to study pragmatics or communication is riskier, especially where subjects expect the dialogue to unfold as a sequence of contingent utterances. One possible compromise is to use a live confederate trained to follow preset rules for responding contingently (see, e.g., Brennan, 1991; Horton & Keysar, 1996; Stent et al., 2008). Even then, the resulting dialogue may lack naturalness, particularly in timing (although this may be handled by giving subjects appropriate attributions). In any event, we argue that it is too risky to use *fully* scripted confederates to stage an authentic dialogue, as conversational interaction includes both linguistic and nonlinguistic elements that are too fine (or too poorly understood) to anticipate in advance. Instead, we suggest that, when a confederate is deemed necessary to standardize the situation, *confederate speakers who produce utterances scripted only at a limited number of critical points, modeled after naturally occurring utterances, may be able to respond contingently and convincingly enough at noncritical points to serve as authentic dialogue partners.*

A fourth recommendation involves *initiative*: Collaborative tasks can be chosen such that the task initiative makes it easier to partially script a confederate's behavior. At the points in the dialogue at which the confederate's role is to take the initiative,

such as by asking the subject a question or giving instructions (rather than responding to what the subject says), such utterances are often easier to script convincingly. In contrast, we argue that in studies of language production, which require subjects to take the initiative to speak, *it can be particularly risky to deploy confederates as addressees*, especially when their role is to serve as an audience to the same utterances or stories retold over and over again by different subjects. The results from storytelling experiments that have focused specifically on speakers' adaptations to addressees' knowledge have established rather conclusively that an addressee who knows too much or who behaves in unexpected ways can affect what the speaker says (whether through the speaker's explicit awareness and expectations or through the addressee's implicit feedback cues; Bavelas et al., 2000; Galati & Brennan, 2006, 2010; Kuhlen, 2010; Kuhlen & Brennan, 2010; Lockridge & Brennan, 2002). Another source of risk in the decision to use a confederate in the addressee role is that addressee behavior is rarely clearly defined, specified, or even understood by the experimenter. Studies using confederate addressees often do not even report what kind of instruction or training was given to the confederate in preparation for their role.

This recommendation about task initiative interacts with the recommendation about giving confederates the right kind of knowledge to support their role in the experimental task. It appears to be less risky for a confederate addressee to participate in a task in which he or she has real informational needs (e.g., as matcher in a referential communication task; see Brennan & Schober, 2001). Rather than excluding confederates entirely from the role of addressee, one solution would be to develop experimental tasks that engage confederate addressees by giving them such needs. In fact, this is advisable whether the addressees are to be confederates or naive subjects. If addressees not only listen, but also collaborate with the speaker on a task with genuine goals, they are more likely to be engaged, as well as to behave authentically.

As a fifth recommendation, we advocate that research involving confederates should *report detailed information on how confederates were integrated into the experimental protocol*. This should include the extent to which confederates knew which condition they were participating in, or what insight they might have had into the experimental hypotheses. Also, it should be reported how often a confederate participated in the experiment and, as a result, how experienced they may have become with the conversational task (experience with the conversational task may be more or less of a concern, depending on whether the confederates had informational needs even after repeated participation). In addition, reports should reveal details of any scripting, as well as any other training or instruction given to confederates. After an experimental session, it is desirable to have confederates note any errors or difficulties in sticking to any scripted portions, or their impressions about any unusual

aspects of the interaction; also, after the experiment is concluded, it is useful to know what confederates thought that their role in the experiment was, as well as how their understanding of the task may have changed over the course of their participation. A summary of such information in the research report may help readers understand and evaluate the results and contributions of the experiment.

Finally, we recommend that language experimenters be deliberate and clear in making trade-offs about whether and how to deploy confederates. *Confederates should probably not be used simply as a convenience*. A carefully chosen dialogue task (such as those deployed by Hanna & Brennan, 2007; Ito & Speer, 2006; Kraljic & Brennan, 2005; or Schafer, Speer, Warren, & White, 2000) may succeed in limiting variability without using a confederate at all; subjects in the speaking role may be led to produce less variable utterances within a dialogue by incidentally exposing them to the desired forms of utterances in a set of practice trials prior to the experiment (e.g., Hanna & Brennan, 2007) or by explicitly training them to use particular expressions or templates (e.g., Schafer et al., 2000). Such approaches vary in how successful they are at eliciting analyzable data (e.g., from Kraljic & Brennan's 50 % success rate in using graphical schematics to elicit reduced and full relative clauses on critical trials, to Hanna & Brennan's 99 % success rate in eliciting referring expressions such as "the red square" in critical utterances). For more discussion of tasks that succeed in eliciting natural dialogue, see Ito and Speer (2006).

Implications and conclusions

Our goal in this article has been to evaluate the use of confederates in studies of language processing in dialogue contexts, and to consider whether and when a confederate can safely replace a naive speaker or addressee. The research question, the level of analysis, and most importantly, the role of the confederate within the experimental task all need to be taken into account. We approached this issue by comparing selected experimental studies that were based on similar questions and similar experimental protocols—with the exception of whether or how they employed confederates—and found that the results were divergent. Although the use of confederates may not be the only reason for this, it does raise the possibility that confederates' behavior may differ from spontaneous or naive subjects' behavior, affecting the interaction in subtle but significant ways. We now conclude with some implications, both for dialogue studies and for studies of social interaction more generally.

Implications for dialogue studies How a confederate has been integrated into an experiment has depended in part on how researchers have conceptualized the nature of

dialogue itself. Although most do acknowledge that conversational partners influence how an interaction unfolds, they may not acknowledge the full extent of this influence. Often the influence is viewed as static and unidirectional—for example, as one conversational partner “priming” the other (Pickering & Garrod, 2004). According to that view, a confederate just needs to pass something like a Turing test (Turing, 1950): If subjects cannot guess that their partner is a confederate, then the confederate passes. The common practice that has followed from this view is that the confederate’s status has been kept covert. While this may be necessary in some sorts of social psychology experiments, especially those that require deceit in order to believably stage an unusual or nonnormative social situation, in most language studies keeping a confederate’s status hidden should not be an end in itself, especially when the goal is to model natural interaction. In fact, some who have gone to great lengths to conceal confederates’ status in dialogue experiments have suggested that the subjects actually behaved no differently, whether they believed that they were interacting with confederates or with other naive subjects (Barr, 2008; Barr & Keysar, 2002; Kronmüller & Barr, 2007). Most importantly, even if an experiment succeeds in concealing a confederate’s status, the confederate’s behavior may differ sufficiently from that of a naive partner to *implicitly* shape the interaction in undesirable ways. If dialogue is understood as a collaborative activity in which partners shape each other’s behavior, representations, and processing in a dynamic, incremental, and reciprocal fashion, then the more serious concern is whether a confederate will know too much and behave accordingly.

Although for some experimental paradigms the use of confederates seems unavoidable, many studies have been able to avoid the use of confederates entirely. As we have discussed, some dialogue studies have focused primarily on the collaboration and coordination between partners (and so, on principle, have not used confederates), and others have succeeded in examining precise constructions or timing in spontaneous production or comprehension without resorting to confederates. This can be accomplished by having pairs of naive subjects collaborate on a referential communication task to match or place objects in a display; the initial and target positioning of the objects is predetermined by the experimenter, creating conditions that allow for the examination of lexical competitor effects or syntactic ambiguity by monitoring the eye movements of one of the conversational partners. Carefully crafted visual stimuli can guide speakers to spontaneously produce utterances with systematic points of disambiguation (Hanna & Brennan, 2007; Haywood et al., 2005; Kraljic & Brennan, 2005). Although some trials will be lost to “exuberant responding” (Bock, 1996), this approach can yield a sufficient number of comparable utterances, such that a standard within-subjects

experimental design can be achieved post hoc without scripting one of the partner’s utterances.

Broader implications for psychological research The use of confederates is alive and well in psychology (and, in fact, seems mandatory for some topics). Previously published studies have had confederates making incorrect judgments (e.g., Ost, Ghonouie, Cook, & Vrij, 2008), falsely recalling items (Roediger et al., 2001), mimicking the subject’s behavior (e.g., Leander, Chartrand, & Wood, 2011), acting cooperatively or competitively (e.g., Hommel, Colzato, & van den Wildenberg, 2009), playing a game fairly or unfairly (e.g., Singer et al., 2006), eliciting a particular mood (e.g., Barsade, 2002), or angering the subject by showing up late to the experiment (e.g., Miles, Griffiths, Richardson, & Macrae, 2010). In the majority of such studies, keeping the confederate’s status covert may legitimately be the main concern.⁵

However, some of the other concerns that we have raised about confederates probably merit broader attention beyond the domain of dialogue studies, especially when the intention is to model naturally unfolding social processes. These concerns apply specifically to confederates whose behavior is highly standardized through predefined scripts. Many phenomena emerge through adaptive co-regulations during social interaction (e.g., De Jaegher et al., 2010; Riley, Richardson, Shockley, & Ramenzoni, 2011; Semin & Cacioppo, 2008; Semin, Garrido, & Palma, 2012). When the goal is to study how ordinary cognitive processes adapt to and coordinate in social interaction, it is safer to avoid using confederates; the relevant unit of analysis to capture these processes is the interaction, not the individual. To the extent that a confederate’s behavior is staged independently of the subject’s behavior and is reduced to serving merely as a stimulus, using confederates may mute, distort, or fail to fully capture the phenomenon under investigation. Moreover, experimental manipulations achieved through the use of confederates are by no means strictly unidirectional, as is exemplified by the fact that confederates themselves assimilate their privately held beliefs to the position that they take at the request of the experimenter (e.g., Laurens & Moscovici, 2005).

A promising approach, for the study of both language in dialogue and other topics, is to develop experimental

⁵ Another issue arises in social psychological research when one confederate is used to represent a token member of a minority group, ingroup, or outgroup (e.g., Gino, Ayal, & Ariely, 2009; Lakin, Chartrand, & Arkin, 2008; Yabar & Hess, 2007). The *Handbook of Social Psychology* (Fiske, Gilbert, & Lindzey, 2010) specifically cautions against this practice; any resulting reactions may be due to the individual’s characteristics rather than being generalizable to the group of which the individual is a member. This problem may be alleviated if more than one person serves as a confederate; even then, researchers are advised to include the confederate as an additional variable into the statistical analysis (Fiske et al., 2010).

manipulations that enable naive partners to fill roles formerly filled by confederates. These manipulations may still involve some deceit, but they allow all partners to act authentically in the moment. For example, rather than instructing a confederate addressee to act inattentively (Jacobs & Garnham, 2007; Pasupathi, Stallworth, & Murdoch, 1998), naive addressees can be secretly preoccupied with a second task (Bavelas et al., 2000; Kuhlen & Brennan, 2010). Even conformity to obviously incorrect perceptual judgments, as famously manipulated through the use of confederates in Asch's (1955) experiment, may be triggered by means of manipulating the perceptions of naive interacting partners by having them surreptitiously view different stimuli than those viewed by the subjects being tested (Mori & Arai, 2010). All of these approaches represent promising alternatives to using confederates while maintaining sufficient experimental control.

Nevertheless, confederates are likely to continue to be widely employed, often for good reasons. To counter concerns about their use, it is essential to have a nuanced understanding of what it means to be an interacting partner. Confederates need to be recognized as being active and potentially influential participants in the interaction, whose behavior needs to be systematically understood, managed, modeled, and monitored. Whether the confederate acts as speaker, addressee, or both, and *especially* when the goal is to investigate partner-specific adaptation in language use, it is essential that confederates behave in ways that are consistent with the roles they play in interactions.

Author note This research was supported by the National Science Foundation under Grant Nos. ITR-0325188 and HSD-0527585. We thank many colleagues for helpful discussion, especially members of the Stony Brook Gesture Focus Group, and in particular, Alexia Galati.

References

- Akmajian, A., Demers, R. A., & Harnish, R. M. (1987). *Linguistics: An introduction to language and communication* (2nd ed.). Cambridge: MIT Press.
- Allport, G. W. (1954). The historical background of modern social psychology. In G. Lindzey (Ed.), *Handbook of social psychology, vol. 1* (1st ed., pp. 3–56). Cambridge: Addison-Wesley.
- Allport, G. W. (1969). The historical background of modern social psychology. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology, vol. 1* (2nd ed., pp. 1–80). Reading: Addison-Wesley.
- Allport, G. W. (1985). The historical background of modern social psychology. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology, vol. 1* (3rd ed., Vol. 1, pp. 1–46). New York: Random House.
- Arnold, J. E., Tanenhaus, M. K., Altmann, R. J., & Fagnano, M. (2004). The old and thee, uh, new: Disfluency and reference resolution. *Psychological Science, 15*, 578–582. doi:10.1111/j.0956-7976.2004.00723.x
- Asch, S. E. (1955). Opinions and social pressure. *Scientific American, 193*, 31–35. doi:10.1038/scientificamerican1155-31
- Banaji, M. R., & Crowder, R. G. (1989). The bankruptcy of everyday memory. *American Psychologist, 44*, 1185–1193. doi:10.1037/0003-066X.44.9.1185
- Bandura, A., Ross, D., & Ross, S. A. (1961). Transmission of aggression through imitation of aggressive models. *Journal of Abnormal and Social Psychology, 63*, 575–582.
- Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science, 15*, 415–419. doi:10.1111/j.0956-7976.2004.00694.x
- Bard, E. A., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language, 42*, 1–22.
- Barr, D. J. (2008). Pragmatic expectations and linguistic evidence: Listeners anticipate but do not integrate common ground. *Cognition, 109*, 18–40.
- Barr, D. J., & Keysar, B. (2002). Anchoring comprehension in linguistic precedents. *Journal of Memory and Language, 46*, 391–418.
- Barr, D. J., & Keysar, B. (2005). Mindreading in an exotic case: The normal adult human. In B. F. Malle & S. D. Hodges (Eds.), *Other minds: How humans bridge the divide between self and other* (pp. 271–283). New York: Guilford Press.
- Barsade, S. G. (2002). The ripple effect: Emotional contagion and its influence on group behavior. *Administrative Science Quarterly, 47*, 644–675.
- Bavelas, J. B. (2005). The two solitudes: Reconciling social psychology and language and social interaction. In K. L. Fitch & R. E. Sanders (Eds.), *Handbook of language and social interaction* (pp. 179–200). Mahwah: Erlbaum.
- Bavelas, J. B., & Chovil, N. (2000). Visible acts of meaning: An integrated message model of language in face-to-face dialogue. *Journal of Language and Social Psychology, 19*, 163–194.
- Bavelas, J. B., & Chovil, N. (2006). Hand gestures and facial displays as part of language use in face-to-face dialogue. In V. Manusov & M. L. Patterson (Eds.), *Handbook of nonverbal communication* (pp. 97–115). Thousand Oaks: Sage.
- Bavelas, J. B., & Coates, L. (1992). How do we account for the mindfulness of face-to-face dialogue? *Communication Monographs, 59*, 301–305.
- Bavelas, J. B., Coates, L., & Johnson, T. (2000). Listeners as co-narrators. *Journal of Personality and Social Psychology, 79*, 941–952.
- Bavelas, J. B., Gerwing, J., Sutton, C., & Prevost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language, 58*, 495–520.
- Bell, A. (1984). Language style as audience design. In N. Coupland & A. Jaworski (Eds.), *Sociolinguistics: A reader and coursebook* (pp. 240–250). New York: St Martin's Press.
- Bergmann, J. R. (1981). Ethnomethodologische Konversationsanalyse. In P. Schröder & H. Steger (Eds.), *Dialogforschung: Jahrbuch 1980 des Instituts für deutsche Sprache* (pp. 9–51). Düsseldorf: Schwann.
- Bless, H., Strack, F., & Schwarz, N. (1993). The informative functions of research procedures: Bias and the logic of conversation. *European Journal of Social Psychology, 23*, 149–165.
- Bock, J. K. (1996). Language production: Methods and methodologies. *Psychonomic Bulletin & Review, 3*, 395–421.
- Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition, 75*, B13–B25.
- Branigan, H. P., Pickering, M. J., McLean, J. F., & Cleland, A. A. (2007). Syntactic alignment and participant role in dialogue. *Cognition, 104*, 163–197. doi:10.1016/j.cognition.2006.05.006
- Brennan, S. E. (1990). Conversation as direct manipulation: An iconoclastic view. In B. Laurel (Ed.), *The art of human-computer interface design* (pp. 393–404). Reading: Addison-Wesley.
- Brennan, S. E. (1991). Conversation with and through computers. *User Modeling and User-Adopted Interaction, 1*, 67–86.

- Brennan, S. E. (2005). How conversation is shaped by visual and spoken evidence. In J. Trueswell & M. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions* (pp. 95–129). Cambridge: MIT Press.
- 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. doi:10.1037/0278-7393.22.6.1482
- Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language*, *44*, 274–296. doi:10.1006/jmla.2000.2753
- Brennan, S. E., & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language*, *34*, 383–398.
- Brown, P. M., & Dell, G. S. (1987). Adapting production to comprehension: The explicit mention of instruments. *Cognitive Psychology*, *19*, 441–472.
- Brown-Schmidt, S. (2009). Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of Memory and Language*, *61*, 171–190. doi:10.1016/j.jml.2009.04.003
- Brown-Schmidt, S. (2012). Beyond common and privileged: Gradient representations of common ground in real-time language use. *Language & Cognitive Processes*, *27*, 62–89.
- Brown-Schmidt, S., Gunlogson, C., & Tanenhaus, M. K. (2008). Addressees distinguish shared from private information when interpreting questions during interactive conversation. *Cognition*, *107*, 1122–1134. doi:10.1016/j.cognition.2007.11.005
- Brown-Schmidt, S., & Tanenhaus, M. K. (2008). Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive Science*, *32*, 643–684.
- Bruehl, D., & Solar, D. (1970). Systematic variation in the clarity of demand characteristics in an experiment employing a confederate. *Psychological Report*, *27*, 55–60.
- Cacioppo, J. T., & Berntson, G. G. (2004). *Essays in social neuroscience*. Cambridge: MIT Press.
- Chafe, W. L. (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge: MIT Press.
- Chomsky, N. (1980). Rules and representations. *The Behavioral and Brain Sciences*, *3*, 1–15. doi:10.1017/S0140525X00001515. disc. 15–61.
- Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
- Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington: American Psychological Association.
- Clark, H. H., & Carlson, T. B. (1982). Hearers and speech acts. *Language*, *58*, 332–373.
- Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, *50*, 62–81. doi:10.1016/j.jml.2003.08.004
- Clark, H. H., & Murphy, G. L. (1983). Audience design in meaning and reference. In J. F. LeNy & W. Kintsch (Eds.), *Language and comprehension* (pp. 287–299). Amsterdam: North-Holland.
- Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as collaborative process. *Cognition*, *22*, 1–39.
- Congdon, S. P., & Schober, M. F. (2002, November). *How examiners' discourse cues affect scores on intelligence tests*. Paper presented at the 43th Annual Meeting of the Psychonomic Society, Kansas City, MO.
- Conrad, F. G., & Schober, M. F. (2000). Clarifying question meaning in a household telephone survey. *Public Opinion Quarterly*, *64*, 1–28.
- Cozby, P. C. (1972). Self-disclosure, reciprocity, and liking. *Sociometry*, *35*, 151–160.
- Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken language comprehension. *Journal of Memory and Language*, *47*, 292–314.
- De Jaeger, H., Di Paolo, E., & Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in Cognitive Sciences*, *14*, 441–447. doi:10.1016/j.tics.2010.06.009
- Deppermann, A. (2008). *Gespräche analysieren: Eine Einführung* (4th ed.). Opladen: Leske & Budrich.
- De Ruiter, J. P., Mitterer, H., & Enfield, N. J. (2006). Predicting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, *82*, 515–535.
- Duncan, S., Kanki, B. G., Mokros, H., & Fiske, D. W. (1984). Pseudounilaterality, simple-rate variables, and other ills to which interaction research is heir. *Journal of Personality and Social Psychology*, *46*, 1335–1348.
- Dutton, D. G., & Aron, A. (1974). Some evidence for heightened sexual attraction under conditions of high anxiety. *Journal of Personality and Social Psychology*, *30*, 510–517.
- Ekeocha, J. O., & Brennan, S. E. (2008). Collaborative recall in face-to-face and electronic groups. *Memory*, *16*, 245–261.
- Ferreira, F., & Swets, B. (2005). The production and comprehension of resumptive pronouns in relative clause “island” contexts. In A. Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones* (pp. 263–278). Mahway: Erlbaum.
- Ferreira, V. S., & Dell, G. S. (2000). The effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, *40*, 296–340.
- Ferreira, V. S., Slevc, L. R., & Rogers, E. S. (2005). How do speakers avoid ambiguous linguistic expressions? *Cognition*, *96*, 263–284. doi:10.1016/j.cognition.2004.09.002
- Fiske, S. T., Gilbert, D. T., & Lindzey, G. (2010). *Handbook of social psychology* (5th ed.). Hoboken: Wiley.
- Forster, K. I., & Masson, M. E. J. (Eds.). (2008). Emerging data analysis [Special Issue]. *Journal of Memory and Language*, *59*(4).
- Fox Tree, J. E. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language*, *34*, 709–738.
- Friedman, N. (1967). *The social nature of psychological research*. New York: Basic Books.
- Frith, U., & Frith, C. (2010). The social brain: Allowing humans to boldly go where no other species has been. *Philosophical Transactions of the Royal Society B*, *365*, 165–176. doi:10.1098/rstb.2009.0160
- Fussell, S. R., & Krauss, R. M. (1989). Understanding friends and strangers: The effects of audience design on message comprehension. *European Journal of Social Psychology*, *19*, 509–525.
- Fussell, S. R., & Krauss, R. M. (1992). Coordination of knowledge in communication: Effects of speakers' assumptions about what others know. *Journal of Personality and Social Psychology*, *62*, 378–391.
- Galati, A., & Brennan, S. E. (2006). Given–new attenuation effects in spoken discourse: For the speaker, or for the addressee? In *Abstracts of the Psychonomic Society, 47th Annual Meeting* (p. 15). Austin, TX: Psychonomic Society Publications.
- Galati, A., & Brennan, S. E. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language*, *62*, 35–51. doi:10.1016/j.jml.2009.09.002
- Garfinkel, H. (1952). *The perception of the other: A study in social order*. Unpublished doctoral dissertation, Harvard University, Cambridge, MA.

- Garfinkel, H. (1963). A conception of, and experiments with “trust” as a condition of stable concerted actions. In O. J. Harvey (Ed.), *Motivation and social interaction, cognitive determinants* (pp. 187–238). New York: Ronald Press.
- Garfinkel, H. (1967). *Studies in ethnomethodology*. Englewood Cliffs: Prentice Hall.
- Gino, F., Ayal, S., & Ariely, D. (2009). Contagion and differentiation in unethical behavior: The effect of one bad apple on the barrel. *Psychological Science, 20*, 393–398.
- Goodwin, C. (1979). The interactive construction of a sentence in natural conversation. In G. Psathas (Ed.), *Everyday language: Studies in ethnomethodology* (pp. 97–121). New York: Halsted Press.
- Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers*. New York: Academic Press.
- Goodwin, M. H. (1985). The serious side of jump rope: Conversational practices and social organization in the frame of play. *Journal of American Folklore, 98*, 315–330.
- Hanna, J. E., & Brennan, S. E. (2007). Speakers’ eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language, 57*, 596–615.
- Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science, 28*, 105–115. doi:10.1016/j.cogsci.2003.10.002
- Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language, 49*, 43–61. doi:10.1016/S0749-596X(03)00022-6
- Hari, R., & Kujala, M. V. (2009). Brain basis of human social interaction: From concepts to brain imaging. *Physiological Reviews, 89*, 453–479.
- Harris, C. B., Paterson, H. M., & Kemp, R. L. (2008). Collaborative recall and collective memory: What happens when we remember together? *Memory, 16*, 213–230.
- Hartsuiker, R. J., Pickering, M. J., & Veltkamp, E. (2004). Is syntax separate or shared between languages? cross-linguistic syntactic priming in Spanish–English bilinguals. *Psychological Science, 15*, 409–414. doi:10.1111/j.0956-7976.2004.00693.x
- Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., & Keysers, C. (2012). Brain-to-brain coupling: A mechanism for creating and sharing a social world. *Trends in Cognitive Sciences, 16*, 114–121. doi:10.1016/j.tics.2011.12.007
- Haywood, S. L., Pickering, M. J., & Branigan, H. P. (2005). Do speakers avoid ambiguity during dialogue? *Psychological Science, 16*, 362–366. doi:10.1111/j.0956-7976.2005.01541.x
- Healey, P. G. T., Purver, M., King, J., Ginzburg, J., & Mills, G. J. (2003). Experimenting with clarification in dialogue. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th Annual Meeting of the Cognitive Science Society* (pp. 539–544). Mahwah: Erlbaum.
- Heritage, J. (1984). *Garfinkel and ethnomethodology*. Cambridge: Polity Press.
- Holler, J., & Wilkin, K. (2009). Communicating common ground: How mutually shared knowledge influences speech and gesture in a narrative task. *Language & Cognitive Processes, 24*, 267–289.
- Hollingshead, A. B. (1998). Retrieval processes in transactive memory systems. *Journal of Personality and Social Psychology, 74*, 659–671.
- Hommel, B., Colzato, L. S., & van den Wildenberg, W. P. M. (2009). How social are task representations? *Psychological Science, 20*, 794–798.
- Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition, 59*, 91–117. doi:10.1016/0010-0277(96)81418-1
- Ito, K., & Speer, S. R. (2006). Using interactive tasks to elicit natural dialogue. In P. Augurzky & D. Lenertova (Eds.), *Methods in empirical prosody research* (pp. 229–257). Berlin: Mouton de Gruyter.
- Jacobs, N., & Garmham, A. (2007). The role of conversational hand gestures in a narrative task. *Journal of Memory and Language, 56*, 291–303.
- Kenny, D. A. (1996). Models of non-independence in dyadic research. *Journal of Social and Personal Relationships, 13*, 279–294.
- Kenny, D. A., & Kashy, D. A. (2010). Dyadic data analysis using multilevel modeling. In J. J. Hox & J. K. Roberts (Eds.), *The handbook of advanced multilevel analysis* (pp. 355–371). New York: Routledge.
- Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. New York: Guilford Press.
- Kerr, N. L., & Tindale, R. S. (2004). Small group decision making and performance. *Annual Review of Psychology, 55*, 623–656.
- Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science, 11*, 32–38. doi:10.1111/1467-9280.00211
- Keysar, B., Barr, D. J., Balin, J. A., & Paek, T. S. (1998). Definite reference and mutual knowledge: Process models of common ground in comprehension. *Journal of Memory and Language, 39*, 1–20.
- Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition, 89*, 25–41. doi:10.1016/S0010-0277(03)00064-7
- Korte, C., Ypma, A., & Toppen, C. (1975). Helpfulness in Dutch society as a function of urbanization and environmental input level. *Journal of Personality and Social Psychology, 32*, 996–1003.
- Kraljic, T., & Brennan, S. E. (2005). Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? *Cognitive Psychology, 50*, 194–231.
- Krauss, R. M. (1987). The role of the listener: Addressee influences on message formulation. *Journal of Language and Social Psychology, 6*, 81–97.
- Kronmüller, E., & Barr, D. J. (2007). Perspective-free pragmatics: Broken precedents and the recovery-from-preemption hypothesis. *Journal of Memory and Language, 56*, 436–455.
- Kuhlen, A. K. (2010). Assessing and accommodating addressees’ needs: The role of speakers’ prior expectations and addressees’ needs. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 71*, 7128.
- Kuhlen, A. K., & Brennan, S. E. (2010). Anticipating distracted addressees: How speakers’ expectations and addressees’ feedback influence storytelling. *Discourse Processes, 47*, 567–587.
- Lakin, J. L., Chartrand, T. L., & Arkin, R. M. (2008). I am too just like you: Nonconscious mimicry as an automatic behavioral response to social exclusion. *Psychological Science, 19*, 816–822. doi:10.1111/j.1467-9280.2008.02162.x
- Laurens, S., & Moscovici, S. (2005). The confederate’s and others’ self-conversion: A neglected phenomenon. *Journal of Social Psychology, 145*, 191–207.
- Leander, N. P., Chartrand, T. L., & Wood, W. (2011). Mind your mannerisms: Behavioral mimicry elicits stereotype conformity. *Journal of Experimental Social Psychology, 47*, 195–201.
- Levinson, S. C. (1983). *Pragmatics*. Cambridge: Cambridge University Press.
- Linell, P. (2005). *The written language bias in linguistics: Its nature, origins and transformations*. London: Routledge.
- Lockridge, C. B., & Brennan, S. E. (2002). Addressees’ needs influence speakers’ early syntactic choices. *Psychonomic Bulletin & Review, 9*, 550–557. doi:10.3758/BF03196312
- Martin, J. D. (1970). Suspicion and the experimental confederate: A study of role and credibility. *Sociometry, 33*, 178–192.
- Martin, J. D. (1973). Pejorative factors in the suspicion of deception: The case of “detection” of confederates. *Journal of Social Psychology, 91*, 155–156.
- Matthews, D., Lieven, E., & Tomasello, M. (2010). What’s in a manner of speaking? children’s sensitivity to partner-specific referential

- precedents. *Developmental Psychology*, 46, 749–760. doi:10.1037/a0019657
- Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Psycholinguistic Research*, 25, 319–344.
- Miles, L., Griffiths, J., Richardson, M., & Macrae, N. (2010). Too late to coordinate: Contextual influences on behavioral synchrony. *European Journal of Social Psychology*, 40, 52–60.
- Milgram, S. (1974). *Obedience to authority: An experimental view*. New York: Harper and Row.
- Montague, P. R., Berns, G. S., Cohen, J. D., McClure, S. M., Pagnoni, G., Dhamala, M., & Fisher, R. E. (2002). Hyperscanning: Simultaneous fMRI during linked social interactions. *NeuroImage*, 16, 1159–1164.
- Mook, D. G. (1983). In defense of external invalidity. *American Psychologist*, 38, 379–387. doi:10.1037/0003-066X.38.4.379
- Mori, K., & Arai, M. (2010). No need to fake it: Reproduction of the Asch experiment without confederates. *International Journal of Psychology*, 45, 390–397.
- Orne, M. T. (1959, September). *The demand characteristics of an experimental design and their implications*. Paper presented at the annual meeting of the American Psychological Association, Cincinnati, OH.
- Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 17, 776–783.
- Orne, M. T. (2002). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *Prevention and Treatment*, 5, n.p.
- Ost, J., Ghonouie, H., Cook, L., & Vrij, A. (2008). The effects of confederate influence and confidence on the accuracy of crime judgements. *Acta Psychologica*, 128, 25–32. doi:10.1016/j.actpsy.2007.09.007
- Page, R. A. (1977). Noise and helping behavior. *Environment and Behavior*, 9, 311–334.
- Pasupathi, M., Stallworth, L. M., & Murdoch, K. (1998). How what we tell becomes what we know: Listener effects on speakers' long-term memory for events. *Discourse Processes*, 26, 1–25. doi:10.1080/01638539809545035
- Pfungst, O. (1907). *Das Pferd des Herrn von Osten (Der kluge Hans) [Mr. van Osten's horse (Clever Hans)]*. Leipzig: Johann Ambrosius Barth.
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *The Behavioral and Brain Sciences*, 27, 169–225.
- Reed, K., Peshkin, M., Hartmann, M. J., Grabowewy, M., Patton, J., & Vishton, P. M. (2006). Haptically linked dyads: Are two motor-control systems better than one? *Psychological Science*, 17, 365–366.
- Richardson, D. C., & Dale, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science*, 29, 1045–1060.
- Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The art of conversation is coordination: Common ground and the coupling of eye movements during dialogue. *Psychological Science*, 18, 407–413.
- Richardson, M. J., Marsh, K. L., & Schmidt, R. C. (2005). Effects of visual and verbal interaction on unintentional interpersonal coordination. *Journal of Experimental Psychology: Human Perception and Performance*, 31, 62–79. doi:10.1037/0096-1523.31.1.62
- Riley, M. A., Richardson, M. J., Shockley, K., & Ramenzoni, V. C. (2011). Interpersonal synergies. *Frontiers in Psychology*, 2, 38. doi:10.3389/fpsyg.2011.00038
- Riley, M. A., & van Orden, G. C. (2005). *Tutorials in contemporary nonlinear methods for the behavioral sciences*. Retrieved March 1, 2012, from www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.jsp
- Roberts, G. K., & Bavelas, J. B. (1996). The communicative dictionary: A collaborative theory of meaning. In J. Stewart (Ed.), *Beyond the symbol model: Reflections on the nature of language* (pp. 139–164). Albany: SUNY Press.
- Roediger, H. L., III, Meade, M. L., & Bergman, E. T. (2001). Social contagion of memory. *Psychonomic Bulletin & Review*, 8, 365–371. doi:10.3758/BF03196174
- Rosenberg, M. J. (1965). When dissonance fails: On eliminating evaluation apprehension from attitude measurement. *Journal of Personality and Social Psychology*, 1, 24–42.
- Rosenthal, R. (1966). *Experimenter effects in behavioral research*. New York: Appleton-Century-Croft.
- Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom. *Urban Review*, 3, 16–20.
- Rosenthal, R., & Jacobson, L. (1992). *Pygmalion in the classroom* (Expanded ed.). New York: Irvington.
- Schafer, A. J., Speer, S. R., Warren, P., & White, S. D. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research*, 29, 169–182.
- Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., Voegeley, K. (in press). Toward a second-person neuroscience. *Behavioral and Brain Sciences*.
- Schober, M. F., & Brennan, S. E. (2003). Processes of interactive spoken discourse: The role of the partner. In A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Eds.), *Handbook of discourse processes* (pp. 123–164). Mahwah: Erlbaum.
- Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, 21, 211–232.
- Schober, M. F., Conrad, F. G., & Fricker, S. S. (2004). Misunderstanding standardized language in research interviews. *Applied Cognitive Psychology*, 18, 169–188.
- Schoonbaert, S., Hartsuiker, R., & Pickering, M. J. (2007). The representation of lexical and syntactic information in bilinguals: Evidence from syntactic priming. *Journal of Memory and Language*, 56, 153–171.
- Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, 10, 70–76.
- Semin, G. R., & Cacioppo, J. T. (2008). Grounding social cognition: Synchronization, entrainment, and coordination. In G. R. Semin & E. R. Smith (Eds.), *Embodied grounding: Social, cognitive, affective, and neuroscientific approaches* (pp. 119–148). New York: Cambridge University Press.
- Semin, G. R., Garrido, M. V., & Palma, T. A. (2012). Socially situated cognition: Recasting social cognition as an emergent phenomenon. In S. T. Fiske & C. N. Macrae (Eds.), *Sage handbook of social cognition* (pp. 138–165). Los Angeles: Sage. doi:10.4135/9781446247631.n8
- Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press.
- Shockley, K., Santana, M. V., & Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance*, 29, 326–332.
- Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., & Frith, C. D. (2006). Emphatic neural responses are modulated by the perceived fairness of others. *Nature*, 439, 466–469. doi:10.1038/nature04271
- Smith, E. R., & Semin, G. R. (2004). Socially situated cognition: Cognition in its social context. *Advances in Experimental Social Psychology*, 36, 53–117.
- Sobel, D. M., & Corriveau, K. H. (2010). Children monitor individuals' expertise for word learning. *Child Development*, 81, 669–679.
- Solano, C. H. (1989). The interactive perspective: Getting the social back into social psychology. In M. Leary (Ed.), *The state of social psychology: Issues, themes, controversies* (pp. 35–44). Newbury Park: Sage.

- Stent, A., Huffman, M. K., & Brennan, S. E. (2008). Adapting speaking after misrecognition: A study of hyperarticulation. *Speech Communication, 50*, 163–178.
- Streeck, J. (2003). The body taken for granted: Lingering dualism in research on social interaction. In P. Glenn, C. D. LeBaron, & J. Mandelbaum (Eds.), *Studies in language and social interaction: In honor of Robert Hopper* (pp. 427–440). Mahwah: Erlbaum.
- Stricker, L. J., Messick, S., & Jackson, D. N. (1967). Suspicion of deception: Implications for conformity research. *Journal of Personality and Social Psychology, 5*, 379–389.
- Stricker, L. J., Messick, S., & Jackson, D. N. (1969). Evaluating deception in psychological research. *Psychological Bulletin, 71*, 343–351. doi:10.1037/h0027232
- Swerts, M., & Kraemer, E. (2005). Audiovisual prosody and feeling of knowing. *Journal of Memory and Language, 53*, 81–94.
- Tanenhaus, M. K., & Brown-Schmidt, S. (2007). Language processing in the natural world. *Philosophical Transactions of the Royal Society B, 363*, 1022–1105. doi:10.1098/rstb.2007.2162
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632–1634. doi:10.1126/science.7777863
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge: Harvard University Press.
- Topál, J., Gergely, G., Miklósi, A., Erdőhegyi, A., & Csibra, G. (2008). Infants' perseverative search errors are induced by pragmatic misinterpretation. *Science, 321*, 1831–1834. doi:10.1126/science.1161437
- Triplett, N. (1897). Social facilitation theory. *The American Journal of Psychology, 9*, 507–533.
- Turing, A. (1950). Computing machinery and intelligence. *Mind, XLIX*, 433–460.
- van Berkum, J. J. A. (2012). The electrophysiology of discourse and conversation. In M. Spivey, K. McRae, & M. F. Joannisse (Eds.), *Cambridge handbook of psycholinguistics* (pp. 589–614). New York: Cambridge University Press.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge: Harvard University Press.
- Wardlow Lane, L., Groisman, M., & Ferreira, V. S. (2006). Don't talk about pink elephants! speaker's control over leaking private information during language production. *Psychological Science, 17*, 273–277. doi:10.1111/j.1467-9280.2006.01697.x
- Weldon, M. S., & Bellinger, K. D. (1997). Collective memory: Collaborative and individual processes in remembering. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 1160–1175. doi:10.1037/0278-7393.23.5.1160
- Wilkes-Gibbs, D. (1986). *Collaborative processes of language use in conversation*. Unpublished doctoral dissertation, Stanford University, Stanford, CA.
- Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Cognition, 31*, 183–194. doi:10.1016/0749-596X(92)90010-U
- Yabar, Y., & Hess, U. (2007). Display of empathy and perception of out-group members. *New Zealand Journal of Psychology, 36*, 42–49.
- Zajonc, R. B. (1965). Social facilitation. *Science, 149*, 269–274.