How the Outside Gets In: Modeling Conversational Permeation

David R. Gibson

Department of Sociology, University of Pennsylvania, Philadelphia, Pennsylvania 19104; email: gibsond@sas.upenn.edu

Key Words
interaction, statistical models, sequence analysis

Abstract
Conversation is incrementally, progressively produced, subject to constraints that ensure linearity (one person speaks at a time) irrespective of the identities, motives, and conversational resources of those present. And yet conversation is also receptive to influence from—or permeation by—external factors, such as attributes, formal status, and relationships. This review summarizes conversation-analytic work on how talk-in-interaction is produced and then evaluates quantitative research on permeation in terms of the realism of its assumptions. Research on rates is found particularly wanting, although the robustness of its results presents a challenge to the claim that the meaning of an action is inextricably tied to its local-sequential context. More theoretically adequate are modeling approaches that focus on transitions, sequences, and the local determinants of discrete events. However, these also frequently make unwarranted assumptions, such as that we can generalize from people who speak to those who do not or that what someone does upon speaking can be considered separately from who speaks in the first place. A solution to the second problem is to model who gets recruited from the ranks of all potential speakers to perform a particular conversational action. The review concludes with directions for future research.
INTRODUCTION

We know two things for certain about conversation. First, talk in conversation is incrementally produced, subject to constraints rooted in a sequential environment that is in constant motion. As a result, the set of conversational options is continuously changing. For example, questions demand immediate answers (Sacks et al. 1974); one has to wait for the right moment to voice a complaint, and that moment quickly passes (Sacks 1995, 1:535–37); and a mistaken interpretation has to be corrected immediately after it is detected if it is to be corrected at all (Schegloff 1992). In almost all informal conversation and in much conversation that occurs in organizational settings, what can be appropriately and intelligibly said at any point in time is constrained by what was said, or otherwise communicated, immediately prior.

Second, although there is theoretical reason to think that conversation is somewhat shielded from external relations and statuses (Goffman 1961, pp. 15–81), conversational behavior is demonstrably impacted by exactly such things, the effects of which can be detected in aggregated measures of behavior (rates and frequencies) that are indifferent to the sequential environment in which particular instances of that behavior occurred. In Goffman’s (1961) terms, encounters are at least “semi-permeable” to external influences. Particularly well studied are the effects of sex. At least in mixed-sex task groups, men speak more than women, men are less tentative than women, and men make more task-related suggestions, whereas women produce more socioemotional remarks (Ridgeway & Smith-Lovin 1999). The effects of authority have also been carefully studied: People in positions of formal authority talk more than others, direct a larger proportion of their remarks to the group, and take longer speaking turns (Bales 1953, Shelly & Troyer 2001, Gibson 2003). And although many of these studies were conducted on experimental task groups, occupational and race effects have been detected in real-life jury deliberations (Rose et al. 2006), sex and generational effects pervade family dinner conversations (Vuchinich 1984), boys are more likely than girls to misbehave in the classroom (McFarland 2001), and people are more direct with friends than with acquaintances in casual conversation (Boxer 1993). Although findings from study to study are not always consistent, overall it seems that microinteraction is suffused with the consequences of relations and attributes that predate the encounter in question.

Obviously, these two truths—the first from conversation analysis and the second largely from expectation states research—do not rest easily together. A proper synthesis is probably not possible at this time; in theoretical presuppositions, goals, and evidentiary standards, the two perspectives are simply too far apart. My objective in this review is, rather, the more modest one of harvesting some of the fruits of conversation analysis for the benefit of a more sophisticated quantitative analysis. In the next section, I lay out the core insights of conversation analysis regarding the linear nature of talk, backchannel (e.g., say “uh-huh” while someone else is speaking) less and use fewer qualifiers (Johnson 1994), whereas both formal and informal leaders talk more than others, direct a larger proportion of their remarks to the group, and take longer speaking turns (Bales 1953, Shelly & Troyer 2001, Gibson 2003). And although many of these studies were conducted on experimental task groups, occupational and race effects have been detected in real-life jury deliberations (Rose et al. 2006), sex and generational effects pervade family dinner conversations (Vuchinich 1984), boys are more likely than girls to misbehave in the classroom (McFarland 2001), and people are more direct with friends than with acquaintances in casual conversation (Boxer 1993). Although findings from study to study are not always consistent, overall it seems that microinteraction is suffused with the consequences of relations and attributes that predate the encounter in question.

Obviously, these two truths—the first from conversation analysis and the second largely from expectation states research—do not rest easily together. A proper synthesis is probably not possible at this time; in theoretical presuppositions, goals, and evidentiary standards, the two perspectives are simply too far apart. My objective in this review is, rather, the more modest one of harvesting some of the fruits of conversation analysis for the benefit of a more sophisticated quantitative analysis. In the next section, I lay out the core insights of conversation analysis regarding the linear nature of talk,
the importance of sequential adjacency (how one thing follows another) that results from linearity, and the continual transformation of conversational possibilities that results from adjacency. Then I describe, in more detail, quantitative research on conversational behavior, focusing on distributions (in particular, of speaking turns) and rates (e.g., of interruption and various speech acts) and the association of rates with sex and formal status. I take this work to task for its unrealistic assumptions about how talk is produced, as a result of which it conflates behavior that is potentially attributable to some preexisting characteristic with behavior that is attributable to the sequential environment. Then I outline a quantitative approach that builds on ideas from conversation analysis regarding the sequential production of talk—in the very least, as a way of controlling for local-sequential constraints so as to better discern the effects of structural and attribute variables. After identifying some of the obstacles to such an approach, I review a handful of quantitative studies that have attempted to take the sequential nature of talk seriously, scrutinizing each for the realism of its assumptions. Finally, I identify avenues and challenges for future work.

Before continuing, I offer three caveats. First, this is a review neither of conversation analysis nor of any of the subfields responsible for quantitative research on conversation, in particular the expectation states perspective. Conversation analysts would no doubt present their work differently (e.g., Goodwin & Heritage 1990), especially given their aversion to abstract statements in favor of letting their view of the world surface in the course of empirical analyses of concrete conversational phenomena. There is, however, a coherent theoretical vision, notwithstanding conversation analysts’ claims that they are engaged in “unmotivated” analysis (e.g., Sacks 1984, p. 27; Schegloff 1996a, p. 172).

The starting point is the linear nature of talk. Predominantly, one person speaks at a time (Sacks et al. 1974), at least in the (mainly Western) cultures that conversation analysts have studied. Experimental research supports the observation (Dabbs et al. 1987), but more important from a conversation-analytic perspective is evidence that interlocutors are attuned to the normative nature of linearity and work to uphold it. Much overlapping talk, in fact, involves an aspiring next speaker beginning his or her turn slightly before the current speaker has reached a point of possible completion, so as to lay an early claim on the next turn (Jefferson 1984a, Schegloff 1987b). Although this results in a brief moment of overlap, it be-speaks an orientation to the one-speaker rule as an operative constraint rather than an attempt to circumvent it. And when two people begin speaking at the same time, as sometimes happens, one will generally desist in short order

**CONVERSATION ANALYSIS: THE SEQUENTIAL PRODUCTION OF POSSIBILITY**

Unlike expectation states researchers, who regularly issue updated theoretical formulations (e.g., Webster & Hysom 1998, Simpson & Walker 2002, Berger & Fisek 2006), conversation analysts generally eschew abstract statements in favor of letting their view of the world surface in the course of empirical analyses of concrete conversational phenomena. There is, however, a coherent theoretical vision, notwithstanding conversation analysts’ claims that they are engaged in “unmotivated” analysis (e.g., Sacks 1984, p. 27; Schegloff 1996a, p. 172).

The starting point is the linear nature of talk. Predominantly, one person speaks at a time (Sacks et al. 1974), at least in the (mainly Western) cultures that conversation analysts have studied. Experimental research supports the observation (Dabbs et al. 1987), but more important from a conversation-analytic perspective is evidence that interlocutors are attuned to the normative nature of linearity and work to uphold it. Much overlapping talk, in fact, involves an aspiring next speaker beginning his or her turn slightly before the current speaker has reached a point of possible completion, so as to lay an early claim on the next turn (Jefferson 1984a, Schegloff 1987b). Although this results in a brief moment of overlap, it be-speaks an orientation to the one-speaker rule as an operative constraint rather than an attempt to circumvent it. And when two people begin speaking at the same time, as sometimes happens, one will generally desist in short order.
Finally, when interlocutors find themselves in a genuine fight for the floor, they have techniques for signaling their respective levels of determination to prevail, which will generally resolve the issue in favor of the more determined party (Schegloff 2000).3

Just as important as the one-speaker rule is the fact that a speaker only articulates one word at a time. Although Schegloff (1996b) is skeptical of psycholinguistic research on how preconceived messages are linearized into utterances (especially Levelt 1989), this does not change the fact that speakers must, by some means, decide what to say of the various things that they could say, given their conversational repertoire (knowledge of the other person, of common acquaintances, of world events, etc.), and decide quickly, since to hesitate is to invite someone else to speak first.

The combined result of the one-speaker rule and the one-word-at-a-time constraint is that even in a conversation involving several people most conversational time is filled with a single stream of talk. The organizational challenge, then, is to provide for the dual selection of speaker and his or her content (or, as I say later, his or her “actions”). In a scripted performance, the order of speakers and the content of their utterances are both set in advance, but in informal conversation and much of that which occurs in institutional settings, long stretches of conversation have to be produced according to selections made on the fly. The most parsimonious solution to this problem—and perhaps the only practical one—is to tie both speaker selection and content selection to the immediately prior speaking turn, and particularly its final portion, for if only one thing happened at the tail end of the prior turn, the range of things that can happen next is greatly reduced if the latter is in some way a function of the former.

3Of course, this does not mean that in any one setting only one conversation takes place at a time. Even in a nominally focused gathering like a family dinner, parallel conversations frequently take place, and that is even more true when people are merely physically copresent, as at a cocktail party or in a train station (Egbert 1997, Ingram & Morris 2007).

From the organizational challenge of ensuring linearity, then, comes the centrality of adjacency as a speaker- and content-selecting principle (Schegloff 1988). Many turn transitions are managed through adjacency pairs, like question-answer and greeting-greeting, that have the property that the first part (the question or initial greeting) creates the expectation that the second part (the answer or return greeting) will be immediately forthcoming from whomever was just addressed (Sacks et al. 1974, Schegloff 2007). Thus, the current speaker can address a question, for instance, to one of several listeners and thereby select that person as the next speaker while simultaneously selecting the type of content that person will produce, at least at the beginning of his or her turn.

Sequential constraints are somewhat looser when no adjacency pair first part is issued, but adjacency (as a general principle) still matters in the form of the general expectation of any turn at talk “that, unless otherwise provided for, it is addressed to what just preceded—to its adjacency prior, or that after which it is ‘next’” (Schegloff 2000, p. 19). That is, every speaking turn is expected to be designed in such a way as to be hearable as a response to whatever was said last, and perhaps to the way that it was delivered paralinguistically and nonvocally (Schegloff 1984). Although this leaves a speaker with some latitude in deciding how to respond (Goffman 1981, Gibson 2000), it still vastly winnows the range of things that a person can say relative to the universe of utterances he or she is capable of. Conversational preferences winnow options further, particularly the preference for agreement, which discourages speakers from bluntly disagreeing with, or otherwise taking a position at variance with the thrust of, whatever was said last (Sacks 1987, Lerner 1996a).4 And all these pressures on content

4The empirical generalizations in this paragraph are of course subject to exceptions. Some are warranted by the sequential context. Were one’s roundabout expression of disagreement misconstrued as true agreement, for instance, a more forceful expression might be called for. Other exceptions are tied to institutional contexts. We would, for instance, not expect a
selection go some way toward selecting a next speaker when no adjacency pair first part does the work, for although procedurally that means that the first person to begin speaking gets the floor (Sacks et al. 1974), in practice the advantage will go to whoever can quickly produce an utterance that meets these requirements.

The final step in the argument is to recognize that, because of adjacency, the menu of conversational options is transformed, often abruptly, from one speaking turn to the next, and even from one clause or word to the next as a speaker’s turn unfolds (Gibson 2005a). Thus, a summons demands an immediate response if any response is forthcoming (Schegloff 1968); misunderstanding needs to be repaired in the turn immediately after it is detected or left unremarked upon (Schegloff 1992); and confirmation of a coparticipant’s interpretation of one’s own oblique talk belongs in the turn immediately after that interpretation is offered (Schegloff 1996a). The general point is that “as the utterances change, what’s admissible [next] changes” (Sacks 1995, 2:556). This poses a significant modeling challenge, one that most quantitative researchers have sidestepped, and at a cost.

QUANTITATIVE RESEARCH ON DISTRIBUTIONS AND RATES

Whereas conversation analysis developed under the dual influences of Garkinkel’s ethnomethodology and Goffman’s face-to-face interactionism (Goodwin & Heritage 1990), quantitative small group research began with the work of Bales (e.g., 1950a). In a large assortment of (mostly task-oriented) groups of previously unacquainted individuals, Bales noted a strong tendency for the rapid development of status and role differentiation, as evidenced by a number of patterns. First, when people are rank-ordered by the frequency with which they initiate communications [mostly verbal, but nonverbal communications were also counted (Bales 1950b)], they are simultaneously rank-ordered by the frequency with which they are addressed, which is to say that people who speak more are addressed more. Second, top-ranked initiators direct a larger proportion of their remarks to the group as a whole than do others. And third, top-ranked initiators dispense more opinions and information than others, whereas those who speak the least focus on expressions of agreement, disagreement, and requests for information (Bales et al. 1951; Bales 1953; 1999, chapter 8).

Bales took these patterns to reflect a process whereby individuals judged more capable of contributing to the group’s task are granted more speaking time and positive reinforcement. He also wondered, without any particular theoretical justification, whether the distribution of initiations in a given group might be mathematically harmonic (Bales et al. 1951). The answer was no, and alternative distributions were subsequently proposed and tested (Stephan & Mishler 1952, Horvath 1965, Burke 1974, Skvoretz 1988). These strands of Bales’s thinking were not properly tied together, however, until expectation states theorists formalized Bales’s ideas about status in task groups and predicted that one’s share of speaking turns would be proportional to one’s relative status in the group. The latter is supposedly determined by one’s preexisting characteristics (including diffuse characteristics such as sex, the effects of which depend on cultural stereotypes, and specific characteristics that point to task-relevant expertise), the characteristics of other group members, and endogenous interactional processes between people identical in terms of those same characteristics (Skvoretz 1981, Fisek et al. 1991).

A strength of the idea that one’s speaking turn allocation is a reflection of one’s relative status is that both variables are normalized to sum to unity, so long as the former is calculated as a proportion of all speaking turns. This is sensible given that the one-speaker rule means that
one person’s speech largely precludes anyone else’s, so that individuals’ respective turn allocations cannot vary independently. However, tests of distributional predictions have tended to be less than conclusive (Kadane 1965; Burke 1974; Skvoretz 1988; Robinson & Balkwell 1995, note 7). One reason is a reliance on statistical tests that assume (as the null hypothesis) that the model is accurate but that have little power to decide otherwise given the small number of groups researchers generally have data on. Another reason is that the rejection of one model provides no guidance as to what to put in its place.

After failing to validate the specific distributional model proposed by Skvoretz (1981), Smith-Lovin et al. (1986) note that the model is correct in the ordinal sense that men do, in fact, participate more than women. In fact, most researchers have skipped distributional predictions in favor of examining such associations, testing these using standard statistical methods such as ANOVA or some form of regression. Such work usually calculates dependent variables as rates, most often calculated as the number of times ego engaged in the behavior of interest divided by the amount of time ego spoke, although sometimes the denominator is total group interaction time. Thus, Kollock et al. (1985) find that the more powerful member of a romantic couple (the person credited with greater influence over everyday decisions) is more apt to interrupt but less likely to backchannel than the less powerful member (regardless of sex); Silver et al. (2000) find that managers speak more than nonmanagers (even after undergoing teamwork training); Rose et al. (2006) find that whites speak more on juries than nonwhites; and Johnson (1994) reports that subjects assigned to positions of formal authority engage in fewer positive interruptions (in which the interrupter uses the opportunity to lend support to what was being said), use less noninterruptive overlapping talk, backchannel less, and also use fewer qualifiers than subjects assigned to be subordinates. Dovidio et al. (1988) added an interesting twist, putting men and women together in groups charged with discussing, in randomized order, feminine, masculine, and gender-neutral topics; they report that male dominance in speaking turns, total speaking time, and assorted nonverbal behaviors is overturned when a feminine topic (sewing) is taken up. And a half-century ago, Strodtbeck (Strodbeck & Mann 1956, Strodbeck et al. 1957) found, comparing simple percentages, that business owners spoke more than clerical workers and laborers on mock juries and that men talked more than women and offered more suggestions and opinions, whereas women did more emotional work.

Despite the wealth of statistically significant findings regarding the effects of, in particular, sex and status on conversational behaviors, all this work is subject to a fundamental criticism from the conversation analysis corner. Using laughter as an example, Schegloff (1993) observes that although laughter per minute is a convenient dependent variable for statistical analysis, people do not, in fact, “laugh per minute,” but rather laugh only in particular sequential environments. Laughter per minute, from a conversation-analytic perspective, is merely an artifact of the underlying conversational machinery, a machinery that manufactures opportunities for action at one moment out of whatever transpired the moment before. The same goes for interruptions per minute, backchannels per minute, speaking turns per hour, and the other outcomes described in the last paragraph; in each case, to explain fully any given instance of a behavior, one has to understand its sequential context. The criticism also applies, in modified form, to research that takes the outcome to be frequencies (McFarland 2001), averages (Shelly & Troyer 2001, Shelly et al. 1999), and probabilities so long as no sequential effects are considered (Perrin 2005).

One irony here is that expectation states theorists actually have a very explicit account of interaction sequences. Simplifying Bales (1953; 1999, p. 168), Fisek et al. (1991) speculate that, in a task setting, the following sequence is fundamental and, presumably, fairly ubiquitous:

364 Gibson
First, one person directs to another an action opportunity, such as a question; second, the recipient of the action opportunity produces a performance output, such as a suggestion or expression of an opinion; and third, someone evaluates that output positively or negatively. Fisek et al.’s actual empirical analyses, however, are limited to tests of their predictions about turn distributions, losing both the sequential nature of the opportunity-output-evaluation model and the content of talk.

Shelly is among the few who seem to be concerned about this gap between theoretical story and empirical tests (see also Robinson & Balkwell 1995). In one paper, Shelly (1997) reduces the story to a set of hypotheses about the association between an individual’s involvement in one sort of act and his or her involvement in another. He hypothesizes, for instance, a positive correlation between the number of action opportunities a person produces and the number of performance outputs he or she is the recipient (addressee) of, and finds support for this in experimental groups. Other hypotheses do not fare as well, however, and there is a surprising paucity of evaluations of either sort (positive or negative). But more important for our purposes is the fact that in order to turn the individual into the unit of analysis, Shelly dispenses with sequential analysis as such—so that we do not know, for instance, whether someone high in both action opportunities produced and performance outputs received was actually involved in many question-answer sequences (though obviously that is a reasonable conjecture). In a later paper, however, Shelly & Troyer (2001) estimate a logit model to predict the occurrence of a performance output or action opportunity given the prior act, thereby retaining information about the actual sequences until the very end. Although their findings do not entirely support Fisek et al. (1991), this approaches the sort of research that I advocate, and thus I have more to say about their study below.

The conversation-analytic critique of research on rates is consequential for a number of reasons. The most obvious is this: If what a person says and does in conversation is largely a product of the sequential environments that she finds herself in, analyses that aggregate across those environments will conflate what people do do with what they must do, and what they do not do with what they cannot do. A related problem is that such analyses equate behaviors that are similar when taken out of sequential context but much different when returned to it, such as changing the topic after a lull, which is a cooperative behavior, and changing it when it is still being energetically developed by others, which is arguably a type of power play. Yet the quantitative findings described above do not simply slink away under the cold light of the conversation-analytic critique, particularly when the observed patterns are theoretically predictable (e.g., Berger et al. 1972). Somehow, we have to admit that both attribute effects and sequential effects are at work. And more precisely, we need a way to study the former as they are mediated by the latter.

THE CHALLENGE: MODELING PERMEATION

The challenge for quantitative microsociologists is to model external effects (of things like status and sex) in such a way as to remain true to how conversation is sequentially and incrementally produced. Put differently, the challenge is to model how and when the constraints that operate in conversation to maintain linearity allow in effects from the outside. Drawing on Goffman’s (1961) idea of the semi-permeability of face-to-face encounters, I refer to this as permeation. In general terms, modeling permeation will involve capturing aspects of the sequential environment vis-à-vis the production of any given speaking turn, either by incorporating those aspects as independent variables or by factoring the dependent variable into its sequentially defined constituents (for instance, so as to distinguish between the dependent variable of speaking after being addressed from that of speaking after someone else was addressed).

This is worth undertaking for a number of reasons. First, if not all conversational openings...
are created equal, parameter estimates for external effects that do not control for sequential constraints will be both substantively misleading and, frequently, statistically nonsignificant. Consider the act of talking. No one doubts that a person is, for the most part, more likely to talk after being addressed with a question than after someone else is addressed with a question. Summing across both sorts of sequential environments conflates talk that is sequentially invited with talk that has to be introduced under less favorable circumstances, resulting in a misleading parameter estimate—deflated vis-à-vis the probability of ego speaking after being asked a question, inflated vis-à-vis virtually everything else—and a large standard error on the basis of which an effect might be declared nonsignificant whereas it is really merely contingent.

Second, insofar as researchers are concerned with what actually gets said in a small group, an understanding of sequential constraints is critical in itself. Expectation states theorists, in particular, have a story about how status differences arise in the course of interaction when not determined by prior characteristics: Someone offers a performance output, someone else offers an evaluation, observers infer an incipient status relationship, and the process repeats with different pairs of individuals but building on that original kernel in such a way as to ensure a transitive hierarchy (Fisek et al. 1991, Fararo et al. 1994, Skvoretz & Fararo 1996; see also Chase 1982). This is not entirely implausible, but from moment to moment other constraints apply, such as to respond to whatever was said last and to temper disagreement. It is not difficult to see the potential relevance of such constraints for a developing hierarchy if the constituent relationships do indeed hinge on conversational exchanges lasting a few seconds (see also Leifer 1988).

Third, one’s involvement in particular kinds of sequences may be consequential for the outcomes of encounters. How regularly one is involved in argument sequences (Goodwin & Goodwin 1990, Vuchinich 1990), question-answer sequences (Shelly 1997), repair sequences (Schegloff 1992), and so forth may matter for reputations, satisfaction (Bales 1953, Strodtebeck et al. 1957), feelings of engagement or disengagement (Goffman 1967, Collins 2004), and subsequent relationships (Gottman et al. 2002). Although here I am mainly concerned with modeling conversational behaviors as outcomes, this is another way in which sequential patterns may be important.

But if conversation analysts have all the answers, why try to salvage quantitative analysis? It is worth pausing to consider the benefits of quantification as such. Although I have drawn on conversation analysis to make a general argument about the sequential nature of talk, conversation analysts generally restrict themselves to the study of procedures, such as those for asking for clarification or communicating disagreement. These procedures, discernible in single episodes of talk (Schegloff 1987a), are akin to the practices people employ in navigating through traffic, subject to constraints inherent to two-dimensional roads and the desire to avoid collisions. Neither knowledge of driving practices (including formal rules) nor knowledge of conversational procedures, however, tells us how people choose from among the options available to them at any instant based on attributes, relationships, and the like. Conversation-analytic research on “institutional talk” (Drew & Heritage 1992) goes somewhat further by linking interactional obligations and entitlements to formal roles, like that of doctor (Maynard 1991) and news interviewer (Heritage & Greatbatch 1991). But this does not nearly exhaust the range of more subtle effects of external factors, such as of friendship ties that may be suppressed on particular occasions. For that, we need quantification and the resulting ability to discern incremental effects, patterns to which there are exceptions, and regularities in the kinds of sequences particular kinds of people find themselves involved in even though such involvement may not be demanded by formal roles.

A quantitative microsociology that incorporates conversation-analytic insights is not a simple matter, however. There is, first of all, the problem of categorizing an utterance as a
member of some speech act category such as command or question, an important step in much quantitative research. Consider the case of questions. One difficulty is that the grammatical form of questions is not always the same. “Why did you do that?” is grammatically interrogative, but “I’d like to know why you did that” is not, but amounts to almost the same thing. A second difficulty is that some actions are performed only over the space of several speaking turns. Consider this hypothetical example:

A: What are you doing tomorrow?
B: Not much. Why?
A: I’m painting my house.
B: Do you need help?
A: Yeah.

This is a request-making sequence, but the request itself is spread out over several turns and in fact is only produced in collaboration with the other party.

A second problem is the difficulty of adequately modeling the lagged effects that are necessarily central to representing the sequential environment. Despite the argument made above for the importance of adjacency, on how what happens at one point in time constrains what can happen next, it is not difficult to see that a first-order Markov model falls short. For one thing, there are many situations in which one can hearably respond to something other than what was said in the previous turn, such as when “insert expansion” (Schegloff 2007) is involved, as in this example:

A: Can you help out?
B: Doing what?
A: Painting.
B: Sure.

B’s “Sure” is a response to “Can you help out?” rather than to “Painting,” though it immediately follows the latter. Further, there are times when the most recent turn prompts a reexamination of earlier talk, such as immediately after a misunderstanding of a much earlier utterance is displayed (Schegloff 1992). Still another difficulty with lagged effects is that it sometimes makes more sense to say that a later event caused an earlier one than vice versa. A question establishing someone’s availability to help (“Are you busy tomorrow?”) that precedes the request itself is one example; another is the gesture that slightly precedes the talk that it helps to punctuate, such as when one points to a door before telling someone to walk through it (Schegloff 1984).

Behind both of these problems—of speech act categorization and lagged effects—is a more general one, namely of identifying, in advance, the relevant features of the sequential environment that constrain the range of options and impart specific meanings to those that remain. From a conversation-analytic perspective, the prospects for this do not look particularly good, especially if one subscribes to Schegloff’s (2007, p. 252) view that although the practices employed in conversation are formally describable—as involving adjacency pairs, repair sequences, delayed disagreement, and so on—a full understanding of how an encounter unfolds is impossible without detailed, qualitative analysis of each episode in all its particulars. And yet we have already reviewed evidence that quantitative research returns interpretable results even when little or no sequential information is taken into account. The question, then, becomes this: What is the nature of the world such that statistical models return significant, interpretable results, despite the fact that quantification means ignoring much of what is distinctive to, and thus potentially decisive in, individual episodes of talk?

There are two possible answers. Take the example of speaking frequency or duration, known to be correlated with status characteristics like sex (Smith-Lovin et al. 1986) and

---

5 There are also techniques by which one may signal that one’s talk is knowingly nonpertinent to immediately preceding talk but only a temporary departure from it, such as whispering behind one’s hand or freezing a gesture relevant to the current course of action for the duration of the side comment (Raymond & Lerner 2007).
authority (Johnson 1994, Shelly et al. 1999, Shelly & Troyer 2001). First, talking as such may have some significance, modified perhaps by simple features of the sequential environment such as having been asked a question, but robust to most other details that might otherwise be taken as contributing to the utterance’s character. Thus, Bales wrote that “[t]o take up time speaking in a small group is to exercise power over the other members for at least the duration of the time taken, regardless of the content” (Bales 1970, p. 76). And Sacks (1995, 1:683) similarly recognized the coveted nature of speaking time: “One wants not merely to occupy the floor, but to have the floor while others listen” (see also Derber 1979). The same might be said of other action and behaviors, such as interrupting if, regardless of the occasion, this points to an underlying power disparity.

The second possible answer to the question of how statistical results are possible is that, within a given setting, the sequential environment surrounding a particular type of behavior is consistent enough that the behavior means the same thing even though that meaning depends integrally on the context. Consider the experimental setting common to expectation states research. This normally involves a well-defined task that does not arouse strong emotions as well as the expectation that group members will share in any rewards resulting from its successful completion. In such a setting, questions may, in fact, be something regularly directed by lower-status individuals to higher-status individuals to elicit the latter’s relatively more highly valued opinions, as expectation states theorists hypothesize (Fisek et al. 1991). And this may be true despite the fact that in the wider world the grammatical form of a question may be the vehicle for doing other sorts of things entirely—such as pressing someone to commit to some course of action or even answering a prior question (“How should I know?”).

Whichever it is, the important point is that conversation does appear to display a kind of order that is statistically discernible. Yet conversation analysts have made a powerful case for the importance of sequential context and the incremental, turn-by-turn production of orderly communication. Thus, again, to reconcile the two perspectives, we need modeling approaches that attend to talk’s incremental production while also allowing for the possibility of permeation from without.

**APPROACHES**

There are a number of ways to go about this. I begin with one formal model that makes distributional predictions, but based on an explicit mechanism for allocating speaking turns one at a time. The other approaches focus on, in turn, transitions from one state to the next, transitions from one event to the next, factors predicting the likelihood of a particular event occurring at a particular moment in time, and how people are recruited to perform sequentially defined actions. Each approach has its limitations, however, and I am particularly concerned with discrepancies between a model’s assumptions and what we know about how talk in interaction is actually produced.

**The Horvath Model**

Decades ago, Horvath (1965) proposed a formal model of how speaking rights are assigned on a turn-by-turn basis, from which the exponential turn distribution claimed by Stephan & Mishler (1952) can be deduced. Horvath conjectured that interlocutors can be rank-ordered by speaking priority. Whenever there is an opportunity for a speaker transition, the top-ranked person first decides whether to speak; if he or she opts to pass, the second-ranked person is presented with the option, and so on. If everyone passes on the opportunity to speak, the option returns to the top-ranked person and the process starts anew, and so on until someone steps forward to say something; by continuously cycling through the group until someone decides to speak, the model ensures that someone will eventually do so. Importantly, Horvath assumed that each person has the same probability of speaking conditional upon
higher-ranked individuals having declined, which means that there is equality in willingness to speak but inequality in the distribution of opportunities.

Horvath (1965) speculated that the rank ordering might reflect verbal latency—how long it takes for someone to begin speaking when given the chance. Skvoretz (1988) proposes several variations on the original model, allowing for external status (sex) effects both on one’s position in the hierarchy of opportunity and on one’s willingness to speak given the chance. The results of Skvoretz’s chi-squared tests—applied to the same mixed-sex groups analyzed in Smith-Lovin et al. (1986)—are inconclusive, however, especially as he admits that the apparent superiority of Horvath’s original model is attributable to the fact that, unlike the competing models, it makes no falsifiable predictions about the effects of sex. Here, however, I want to judge Horvath’s model not by its fit to the data (see also Kadane 1965), having already commented upon the difficulty of testing distributional models, but rather by the realism of its assumptions about how turns are allocated. The model stands out as an early attempt to account for speaking inequalities in a way that is mindful of the basic organizational problem of recruiting one and only one person to speak in a given turn. But the process proposed is clearly far-fetched. One problem is that there is no recognition of the many ways in which what happens in one turn is consequential for what happens in the next, including for who speaks. A second problem is that there is no reasonable sense in which a highly ranked person—whether in terms of verbal latency or status—is forced into the conversational wings upon declining the initial opportunity to speak until such a time as everyone else has similarly declined.

State Transitions: Speech and Silence

Another approach (Cappella 1979, 1980) builds on the work of Jaffe & Feldstein (1970) to model speech-silence transitions in dyadic conversations. In the simpler version of the model, conversation can occupy any of four possible states at a given instant: (a) Both parties are talking, (b) neither party is talking, (c) the first person is talking and the second is not, and (d) the second person is talking but the first is not. Cappella (1979) proposes four models of increasing complexity. The Markov model postulates a first-order stochastic process that takes the probability that the system will occupy a particular state at time t to be a function only of the state it occupied at time t-1. The independent decision model is an attempt to incorporate the possibility of individual consistency across encounters—a kind of external effect—and assumes that each party decides whether to speak at t independently, as a function of the state of the system at t-1, and in accordance with propensities inferred from his or her behavior in one or more previous encounters with other partners. The incremental model is similar to the independent decision model, except that a person’s propensities are assumed to be adjusted upward or downward depending on the partner. Finally, the regulation model relieves the assumption of independence, allowing partners to influence one another, so that each partner’s transitional propensities are adjusted over the course of the encounter in response to the similarly evolving propensities of the other person.

Of most interest from the perspective of permeation is the independent decision model because it attempts to reconcile the process by which talk is sequentially produced with propensities that adhere to particular individuals by virtue of personality, cognitive abilities, and the like. Cappella (1979) acknowledges that the assumption of complete between-encounter consistency of behavior is unrealistic, however, and sure enough, when the model is tested using data on eight individuals interacting in various combinations, the independent decision model

---

6 This is Cappella’s (1979) four-state description. Six states are possible if we claim to be able to distinguish who actually has the floor when both individuals are vocalizing. Cappella (1980) tests his models using both descriptions, but the results do not differ in any way that matters here.
model does not fare as well as the Markov model (Cappella 1980). The implication is that people are not, in fact, very consistent from one encounter to the next. That seems to challenge the very idea of permeation, but the model’s problematic assumptions about how talk is produced vitiate any substantive conclusions. Specifically, were interlocutors to make independent speak-versus-silence decisions, as postulated by the model, we would see both more mutual silence and more simultaneous talk than we do. Most of the time, speaker transitions occur in an orderly fashion, and between such transitions there are stretches of talk during which control of the floor is uncontested. Conversation analysts have an explanation. Speaker change, they say, is most likely at some point in a speaker’s turn that can be construed as a transition relevance place, normally a grammatical boundary marking the completion of a minimal conversation action, like a question or display of understanding, to which a response can be fashioned (Sacks et al. 1974, Schegloff 1996b; see also Ford & Thompson 1996 on alternative operationalizations). Speaker transitions are least likely to occur during the chunks of single-speaker talk that occupy the space between transition relevance places, known as turn construction units. As a result, for long stretches, the likelihood of speaker transition is small, and when a transition relevance place arrives, everyone knows that speaker transition is a possibility. Further, there is a whole expectational apparatus that allows one to anticipate when a transition relevance place is likely to appear. This allows us to act in a way that appears instantaneously compatible with the other person’s actions—to remain silent when that person is speaking, and to begin speaking when that person is ready to yield the floor. Thus, far from making their speaker-versus-silence decisions independently, interlocutors recognize the differential obligations and entitlements of speakers and listeners, and the way in which the evolving contours of the turn-in-progress variously forestall and invite coordinated speaker transition.

Action Transitions: Speech Act Sequences

I faulted Horvath and Cappella for ignoring the content of talk and the way that this creates obligations and expectations that structure the talk that immediately follows. A standard way to incorporate content is by classifying utterances according to some speech act scheme. Bales (1950a,b) proposed one particularly well-known system for use in task group settings; the 12 Interaction Process Analysis (IPA) categories include asking for an opinion, giving an opinion, agreeing, and tension release (e.g., joking). Expectation states theorists simplified this along the lines explained earlier, into action opportunities, performance outputs, and positive and negative evaluations (e.g., Shelly 1997). Many other schemes exist (D’Andrade & Wish 1985, Gottman & Notarius 2000).

Although speech act sequences can be analyzed in terms of simple probabilities (e.g., Bales 1953), speech act sequences are most flexibly modeled within a loglinear framework (Bakeman & Gottman 1986, pp. 116–17), one benefit of which is that it is easy to include categorical information about speakers so as to judge whether sequential tendencies vary with the attributes of the individuals involved. Manderscheid et al. (1982), for instance, classify speaking turns according to whether they conclude with moves to exercise control, relinquish control, or neither, and then model these sequences using data on eight married couples. They report that the control direction of a given turn-concluding utterance is best modeled as a third-order Markov process, which means that it depends on the control directions of the three turn-concluding utterances that preceded it; they further report that model fit is not significantly improved by taking the speaker’s identity (husband or wife) into consideration. Manderscheid et al. then use a diagrammatic representation of transition probabilities to examine how the interaction system moves between consecutive three-turn “states” and find that the system gravitates toward consecutive control-exercising moves.
and consecutive control-neutral (i.e., neither exercising nor relinquishing control) moves.

Vuchinich (1984) makes similar use of log-linear models in his study of the dinner conversations of 52 families. Focusing on episodes of family conflict, he classifies speaking turns by whether they entail simple negation (“no”), disagreement (where the point of contention is articulated—“no, I am going out tonight”), or indirect negation (when people take discordant positions but do not, in the space of single speaking turns, acknowledge the state of disagreement expressly). Unlike Manderscheid et al. (1982), Vuchinich (1984) finds that the first-order Markov model fits best, which means that the nature of the current oppositional move depends on the nature of the last move but not those further back. Also unlike Manderscheid et al., Vuchinich finds that family members differ in terms of the sorts of sequences they are implicated in and generally along the lines one would expect. For instance, whoever succeeds a turn in which a father uses a simple negation is relatively unlikely to produce another simple negation compared with whoever speaks after a mother who does the same thing; Vuchinich interprets this to mean that “[p]articipants are less likely to challenge the father ‘toe to toe’ without some mitigation than they are to challenge the mother” (p. 229).

As a final example of the use of loglinear models to study speech act sequences, Shelly & Troyer (2001) test the expectation states account of the sequences alleged to be fundamental to task group deliberations by modeling the effects of the speech act at t-1 on the odds of a performance output at t (using a logit specification), using data on 70 three-person task groups. Surprisingly, in light of the purported ubiquity of opportunity-performance-evaluation sequences, a positive evaluation in one turn does more to elicit a performance output in the next than does an action opportunity. More colloquially, praise is more effective at eliciting an opinion than is a question, although this elides a coding peculiarity that I explain shortly. Although the odds of a performance output (as well as the odds of an action opportu-

nity, modeled separately) are also affected by the status of the (second, i.e., focal) speaker, the experimental condition (not described here), and their interaction, there is apparently no significant interaction of status with the content of the prior act. That is, although high-status individuals are more likely to produce performance outputs, controlling for this, they do not respond to sequential conditions (the content of the prior act) differently than do low-status individuals.

Work on speech act transitions raises a number of issues, most of them anticipated earlier. There is, first of all, the matter of categorizing utterances. All the speech act classification schemes are subject to the conversation-analytic criticism that they were formulated with prior theoretical concerns in mind, rather than through the direct analysis of what people actually achieve in conversation and how they achieve it (Schegloff 2007, p. 8). This is obviously true of Manderscheid et al. (1982) and Vuchinich (1984), whose schemes aggressively collapse utterances into a small number of types. [Family researchers go further still, often reducing an utterance to its value on a single positivity-negativity dimension (Gottman et al. 2002, Gottman & Notarius 2000).] It may also be true of the expectation states scheme, which presupposes the primacy of task-related activities and ignores socioemotional work (Ridgeway & Smith-Lovin 1999, Johnson et al. 1996). Nor is Bales’s IPA scheme immune to the criticism, although it does include socioemotional categories, including antagonism and tension release. Compounding the problem further, both expectation states and IPA schemes count nonverbal behavior: Bales (1950b) considered a sigh combined with fingernail biting as an instance of “shows tension,” while expectation states researchers count a long silence as an action opportunity equivalent to asking a question (Shelly 1997, Shelly & Troyer 2001). This raises a host of difficulties. To take the last example, although a silence can elicit a new topic proposal in ordinary talk (Maynard 1980), and perhaps a performance output (such as a suggestion) in a task group

www.annualreviews.org • Modeling Conversational Permeation 371
setting, a question does far more to limit the range of possibilities in the next turn, both in terms of who talks and what they say. Consequently, lumping the two together seems unjustified and may help explain Shelly & Troyer’s (2001) finding that action opportunities do not reliably call forth performance outputs. And as for other forms of nonverbal behavior, such as gestures and sighs, because these lack the semantic content of words, any attempt to categorize them as equivalent to verbal speech acts seems similarly problematic (see also King 2001).

A second issue, discussed above, is that of identifying the proper order, or the number of prior turns that are permitted, by the model, to matter for the current one. I reviewed some conversation-analytic work suggesting that there is no single proper order, that how far back in time one needs to look to explain the constraints operative in the current turn depends on the sort of sequence the current turn is a part of. That kind of contingency would be difficult to model directly, however, and Vuchinich (1984) and Manderscheid et al. (1982) take the standard approach of adding increasingly high orders (lagged terms) until further improvements in model fit become nonsignificant. (Shelly and Troyer only consider order-1 effects.) The problem with such an approach is that at higher orders data become very sparse, and that is particularly true for Vuchinich, who only considers arguments in family dinner conversations that presumably involved long stretches of companionable chitchat and who, perhaps not coincidentally, settles on an order-1 model. Both for lack of data and because the true order is contingent on the sort of sequence in production, such an approach is likely to understate the extent to which the conversational system remembers what happened earlier.

A third issue arises now for the first time, although it is important in the next section as well. When the dependent variable is what someone says or otherwise does (e.g., interrupts) in a speaking turn, it is common practice to take attributes of that person, such as sex and formal status, as independent variables and then to infer from any significant effects what “males do,” “high-status people do,” etc. Thus, Vuchinich (1984, p. 228) interprets his results as suggesting that “[w]hile daughters are more likely to participate in verbal conflict, they are less likely than males to oppose parents.” That inference, however, is not warranted, and the use of these sorts of models to study permeation is compromised as a result. The reason is that we only have data on people who speak, or insofar as they speak, and nothing whatsoever can be inferred about those who do not, or who do not on a particular occasion.7 All Vuchinich can reasonably conclude is that daughters who speak are more likely to engage in verbal conflict than males who speak but are less likely than males who speak to oppose parents. A daughter who never speaks might, in fact, avoid verbal conflict if allowed (or forced) to assume the floor. This problem, in turn, rests on a deeper one, namely that the models are not predictive of who actually speaks. By taking the identity of the second speaker as having already been decided, and then using his or her attributes as predictors of the content of his or her utterance, researchers are able to sidestep the question of who speaks and pretend that nonspeakers would behave in the same way as speakers do were they given the chance. But this invites misleading inferences, as discussed further below.

### Precipitating Events: Interruptions and Topic Changes

One of the lines of quantitative research that explicitly incorporates conversation-analytic insights is that of Smith-Lovin and colleagues on interruptions and topic change. In the first of these articles, Smith-Lovin & Brody (1989) use a series of logit models to study patterns of interruption in the same mixed-sex task groups

---

7This is not a problem for Manderscheid et al. (1982) because they studied dyads (married couples), so that a given person was always either speaking or being responded to. Vuchinich (1984), in contrast, studied families, and though he does not provide information on the sizes, from his article it is apparent that families of four or more were common.
described earlier. First, they model the odds that an interruption [defined as “an intrusion into the internal structure of the speech act, not corresponding to a possible transition place” (p. 428)] was attempted. Second, they model the odds that the attempted interruption was supportive (e.g., the interrupter agreed or finished the first speaker’s thought), negative (e.g., the interrupter disagreed or changed the topic), or neutral; they refer to this as the attempt’s “affective character.” And third, they model the odds that the interruption attempt was successful (as judged by whether it forced the initial speaker to stop speaking prior to a possible transition place). In each case, the independent variables include the sex of the initial speaker and the sex of the second speaker (and possible interrupter). The most memorable findings result from the first and third analyses: Men attempt to interrupt women more regularly than they attempt to interrupt men, whereas women attempt to interrupt men and women equally; and men yield to male interrupters more readily than to female interrupters, whereas women yield to male and female interrupters equally. Taken together, these findings suggest that men care more about sex (as a status characteristic) than women do.

This analysis does not have an explicitly temporal or sequential dimension, except insofar as it attends to the sexes of consecutive speakers (which is better than simply asking whether males interrupt more frequently than females). The distinction between the three types of affective character does, however, reflect a conversation-analytic sensibility, inasmuch as it is an acknowledgment that the meaning of an intrusion depends on the relationship between what the interrupter says and what the initial speaker was in the midst of saying. In another article, Robinson & Smith-Lovin (1990) bring temporality to the fore, using event history methods to consider two temporal (if not precisely sequential) variables: how long the group has been talking and how much time has elapsed since the last interruption. They find that interruptions become more likely with time but that the time since last interruption has no bearing on the likelihood that the current turn will be interrupted, contrary to the idea that interrupting behavior is contagious or that one interruption makes a retaliatory interruption more likely.

Okamoto & Smith-Lovin’s (2001) article on topic change does even more to incorporate conversation-analytic insights into a statistical model. They distinguish between topic changes that are collaborative (when group members cooperatively bring talk on a particular topic to a close), linked (when the topic-changing turn begins with some recognition of the prior talk), and sudden (when there is neither cooperation nor recognition). As with the types of interruption in the earlier paper, this distinction embodies the insight that the causes and import of a topic change may depend on how it relates to immediately preceding talk. They then use an event history model to identify the variables affecting the hazard of a topic change in a given turn. Importantly, those variables include not only the sex of the focal and previous speaker but also several variables that characterize the sequential context: whether the focal turn was preceded by a pause of 0.5 seconds or longer, which may indicate failure of the topic to adequately engage interlocutors (Maynard 1980); number of words in the previous turn, where a short turn may similarly indicate lack of interest; and a dummy variable for whether the person speaking in the focal turn was originally responsible for introducing the topic now at risk of being changed. The last of these variables anticipates the possibility that one’s conversational prerogatives are a function not of prior characteristics but of earlier conversational action. Further, the authors include several variables capturing a person’s participation history: the number of words the focal speaker has said so far in the discussion, the number of words the previous speaker has said so far, the number of times the focal speaker has introduced topic changes, and the number of times the previous speaker has had his or her topic changed by whoever spoke next.

Okamoto & Smith-Lovin find that men are more likely to change the topic after women
speak than after men speak, which is reminiscent of Smith-Lovin & Brody’s (1989) finding about the effect of sex on interruptions. In terms of participation history, talkative speakers are less likely to see their topics changed in the next turn, especially by other talkative speakers, and a new topic is more likely to be introduced by someone who has introduced new topics in the past and following turns by speakers who have frequently seen their topics changed. Finally, the conversation-analytic variables have the predicted effects: A topic change is more likely after a pause, after a short turn, and when the focal speaker was originally responsible for introducing the topic at risk. (Results pertaining to type of interruption are harder to interpret, and I refrain from summarizing them here.)

This work does much to incorporate features of the sequential environment in which interruptions and topic changes do and do not occur. Further, it seeks to explain phenomena that have concerned conversation analysts (e.g., Jefferson 1984b, Schegloff 2002) and that are arguably less vulnerable to the accusation (levied at speech act analyses) that the coding categories were imposed from on high, without concern for the kind of distinctions people demonstrably make as they talk. That said, the final issue identified in connection with the analysis of speech act sequences is relevant here as well. Smith-Lovin & Brody (1989) and Okamoto & Smith-Lovin (2001) take information about the second speaker in the two-turn sequence as potentially predictive of the occurrence or nonoccurrence of an interruption or topic change, as a result of which they can only justifiably draw conclusions about what people do upon speaking, not about who speaks in the first place.

This may seem like splitting hairs, and if men, upon speaking after women, are especially likely to interrupt or change the topic, then a given man is probably more likely than a given woman to speak after a woman speaks and to use the opportunity to change the topic. But it is not difficult to imagine a scenario in which the results of this approach might lead to unwarranted conclusions. Take the case of topic change, the sex effects on which are, in fact, weaker than Okamoto & Smith-Lovin had anticipated. Imagine, for the sake of simplicity, that topics only get changed when the discussion around them begins to flag, and that when discussion starts to flag the topic always get changed. If men are always the ones who rise to the occasion to change the topic, gender will be strongly collinear with the indicators of a flagging topic, and in the extreme case sex may appear to have no significant effect, though, at the stage of speaker selection, it definitely does.

The problem here is that speaker selection is being artificially separated from content action (or, more generally, action selection)—two things that, from a conversation-analytic perspective, go together, as explained above. This invites taking the speaker as the cause of the content, though in many circumstances it is the content—or the opportunity to produce a certain type of content—that selects the speaker, as in the example at the end of the last paragraph. A related problem arises in connection with Smith-Lovin & Brody’s interruption paper. There, they take the focal speaker’s sex as potentially predictive of whether an interruption occurs, but if no interruption occurs, that is because no one interrupted; there is no reason to give special credit to whoever eventually and properly spoke at some transition relevance place. Indeed, in taking the noninterrupter/focal speaker’s sex as a cause of the noninterruption, they are reversing the order of events: In real conversational time, the opportunity to interrupt comes before the opportunity to take the floor noninterruptively, so that attributes of whoever does the latter cannot be considered the cause of no one doing the former, at least in a group of three or more people.

**Recruitment to Act: Participation Shifts**

Recent research on conversational behaviors like interrupting, topic change, and assorted speech acts point to a growing impatience with the study of speaking turn/time allocations as such. After all, people not only speak; they say
and do things when they speak, and perhaps what they say and do and how others respond to them more directly captures the effects of status than how much airtime they consume. The problem, we have seen, is that this has led to models that sever speaker selection from content selection and take speaker selection as already decided, although this means modeling one side of the conversational coin as if the other side were not flipping with it.

Gibson (2003, 2005b) has attempted to get around this problem by studying how people get recruited to speak and to speak in a certain way. Key to his work is the idea of a participation shift, or P-shift. P-shifts refer to the way in which people are shuffled between the statuses of speaker, target (addressee), and undressed recipient (i.e., everyone else) from one turn to the next, or within the space of a given speaking turn. There are, in fact, exactly 13 such P-shifts. An example is the prototypical two-turn exchange involving two people addressing each other in turn; this is denoted as AB-BA, which means that some person A talks to some person B, and then, in the following turn (after the hyphen), B talks to A. This is not the only thing that can happen after one person addresses another, however. B can also talk to the group (AB-B0, where 0 is the group) or to some other individual (AB-BY), and if someone other than B talks, that person can talk to A (AB-XA), to B (AB-XB), to some other individual (AB-XY), or to the group (AB-X0). After someone talks to the group (A0-), the next speaker can address that person (A0-XA), or someone else (A0-XY), or the group (A0-X0). Finally, a speaker can change target in the middle of a turn, so as to address two people in succession (AB-AY), an individual and then the group (AB-A0), or the group and then an individual (A0-AY) (Gibson 2003).

A number of things should be noted about this scheme. First, P-shifts abstract from named individuals to focus on conversational statuses and the movement of individuals between them. This builds on the conversation-analytic idea that obligations and entitlements apply to individuals not (or not only) by virtue of their enduring roles and identities, but by virtue of more ephemeral statuses acquired through their conversational actions and the actions of others (see also Parker 1988). Second, the scheme focuses our attention on consecutive speaking turns and the constraints that one turn places upon the next, consistent with conversation analysts’ idea of adjacency. Third, although a P-shift may implicate up to four people (AB-XY), the shift itself is produced by a single individual: the person speaking in the second of the two turns (or the person changing targets mid-turn). Thus P-shifts are simultaneously a matter of individual action and collective involvement. Fourth, the scheme ignores content, under the assumption that, second only to the question of who speaks, the question of who is addressed is the most fundamental aspect of a turn, as it determines for whose subsequent conduct the content of the turn is primarily consequential (cf. Levinson 1988).

The P-shift framework is at the core of two articles. In one, Gibson (2003) analyzes P-shift frequencies for evidence of the operation of conversational rules (such as an apparent prohibition against addressing anyone other than the previous speaker or previous target) and, controlling for these frequencies, for evidence of permeation by sex and formal status, using data on 13 managerial teams. Sample findings include that superiors are more likely than others to talk to the group, regardless of what happened in the previous turn, and that men are more likely than women to address the group, whereas women are more likely to address particular individuals and then to be addressed in turn.

Because people vary immensely in their propensities to speak in a turn after one in which they were not addressed, Gibson finds it necessary to analyze the effects of sex and formal status on such P-shifts (e.g., AB-X0, A0-XY) by calculating a person’s likelihood of addressing the previous speaker, previous target, or someone else, conditional on that person already having the floor. That amounts to the same splitting of speaker selection and action selection that I criticized above. However, a person’s
tendency to so take the floor is also analyzed, and a person’s tendency to be involved in a P-shift beginning with him or her speaking or being addressed was calculated as the probability of, for instance, ego talking to the group conditional upon just having been addressed—that is, the probability of AB-B0 given the initial AB—which respects the simultaneity of speaker and action (here, target) selection.

In a second article, Gibson (2005b) takes a much different approach to study the effects of networks, such as those of friendship and being coworkers, on P-shift behavior, and in particular on the regularity with which two people are coinvolved in particular P-shifts. He asks whether ego is especially likely to complete a P-shift relative to someone to whom ego is tied in some network, controlling for ego’s overall propensity to complete that P-shift and the sequential constraint that, for instance, ego cannot talk to alter after alter has addressed the group if alter never addresses the group.

A sample finding is that friends and coworkers are especially apt to “piggyback” (Goodwin & Goodwin 1990) by addressing the same third person in turn.

Gibson’s statistical analysis is based on a modified permutation test that cannot be explained in detail here. One noteworthy feature is that the analysis is performed on each P-shift separately, out of respect for the possibility that the meaning of speaking depends on the sequential environment in which it occurs. That said, like other researchers, Gibson can conclude nothing about what nonspeakers would have done had they had the opportunity to speak. However, his approach does better with respect to the problem of artificially separating speaker selection from content or action selection. The reason is that it asks not how a person’s attributes affect what she does upon speaking, but about how her network position affects the likelihood that she will be recruited, from the ranks of all potential speakers, to perform a particular conversational action, such as addressing someone who just spoke to the group. This is the difference between explaining what a speaker decides to say and explaining who finds their way into the position of doing one of the things that can be done at a particular conversational juncture.

**CHALLENGES**

In this section, I identify some challenges and opportunities for the quantitative analysis of conversational sequences in general and the analysis of permeation in particular. The first bears directly upon a problem with all the work reviewed so far on extradyadic encounters (involving three or more people), namely of generalizing from people who speak to those who do not. The second and third challenges relate to different dimensions of long-term temporality. The fourth challenge returns us to the sticky matter of content. Finally, I consider the obvious but problematic fact that, although predominantly one person speaks at a time, if we take nonverbal communication into account it remains the case that interactional behavior occurs concurrently, and perhaps consequentially so.

**Modeling the Individual**

I have argued that models that take the identity of the second speaker in a two-turn sequence as pregiven can produce misleading results, or results that are easily misinterpreted. One solution is to model the behavior of particular individuals, such as by means of an event history model that models the hazard of a particular individual talking or, better yet, making a particular kind of intervention, such as an interruption, topic change, or uncommon P-shift. The problem is that the one-speaker rule will ensure weak effects and high standard errors when the permeating attribute (such as being a male) is shared by many people, most of whom will not be able to act on it on any given occasion even when the sequential environment permits the action. A possible solution is to include in the equation information about the person’s position in the speaking hierarchy (however determined) as an additional predictor. The attribute might then be seen as predictive (if in fact it is
predictive) of one’s motivation to perform some action, whereas the status variable would capture the opportunities one is given to actually do so—reminiscent of Horvath (1965). Such a model would still, however, violate the independence assumption, but in the annals of quantitative microsociology it would hardly be unique in doing so.

**Conversational History**

Most sequential models assume that it is the most recent past that matters for the present; this is even true when higher-order effects are considered because conventional tests rarely favor orders higher than three (Manderscheid et al. 1982). Part of the reason is that data are stretched thin when higher orders are considered (Parker 1988). But another reason, illustrated earlier, is that there is, in fact, no single order equally applicable to all conversational moments, but rather coherent conversational sequences of various lengths (Schegloff 2007). One option might be to use methods tailored to the analysis of entire sequences (Abbott 1995) in an effort to identify typical sequences, such as of argument or apology. Another, more time-consuming option is to conduct detailed analyses of transcripts to identify the type of sequence in progress at each moment and its historical depth, information that could then be used to characterize the focal turn. Another problem is that events early in a conversation may become newly consequential later, even when not ostensiblly so during the interim (e.g., Schegloff 1992, Schegloff & Sacks 1973). One option is to include dummy variables capturing features of the sequential environment established further back in time; an example is Okamoto & Smith-Lovin’s (2001) inclusion of a dummy variable for whether it was the focal speaker who earlier introduced the current topic.

**Conversational Phases and Trends**

I have largely ignored the matter of phases and trends in task groups, including problem-solving phases (Bales & Strodbeck 1951, Poole 1981, Arrow et al. 2004), the process of role differentiation (Bales 1953), and the ways group dynamics change as deadlines approach (Gersick 1989). The reason is that conversation analysis has little to say about such trends.8 The obvious bridging question is how the application of rules and the production of particular sorts of sequences pursuant to those rules change with the passage of time. Some hints are found in Robinson & Smith-Lovin’s (1990) finding that interruptions become more likely as time passes and Okamoto & Smith-Lovin’s (2001) finding (not mentioned earlier) that topic changes become less frequent. A problem, however, is that these models take time as a continuous variable and thus are insensitive to abrupt turning points that might be responsible for what otherwise seem like gradual changes in likelihood. Methods for identifying turning points exist (Griffin & Isaac 1992), although their utility is greatest when the periods to each side of the turning point are internally stationary.

**Content**

Speech act sequence analysis has fallen on hard times. Such analyses tend to deliver results that are either obvious or uninterpretable, in the latter case leading us to become suspicious of the coding rules involved. Perhaps for this reason, sociologists, at least, favor analyzing conversational behaviors such as interrupting and backchanneling. But it remains the case that people say things in conversation and that what they say is consequential for how they are perceived and for the course of subsequent interaction. One possibility is to code not for semantics, but for practices such as finishing someone else’s sentence (Lerner 1996b), repeating oneself (Schegloff 1987b), and responding to a question with a question (Schegloff 1981).

---

8An exception is research on how people manage their talk in anticipation of the work that has to be done in an encounter and the time that remains to do it (e.g., Clayman 1989). One question is whether the trends in question depend on such awareness; in Gersick’s (1989) case, they evidently do.
The problem is that, although such practices may be reliably identified, they are fairly bloodless compared with the actions they advance—such as "disagreeing, offering, contesting, requesting, teasing, finessing, complying, performing, noticing, promising, and so forth" (Schegloff 2007, p. 7)—and there is no one-to-one mapping of one onto the other. Some other way must thus be found to categorize content while avoiding the pitfalls of classic speech act analysis, particularly the imposition of artificial categories and the assumption that speech acts are neatly enclosed within individual utterances or speaking turns.

Above the level of individual speech acts and coloring their interpretation are activities and their associated interpretive frames (Levinson 1992). Consider the simple fact that playing and fighting are different activities and entail different understandings of the constituent actions (Bateson 1972, pp. 177–93). This fact is consequential not only for what people carry away in their heads, but for sequential patterns: A push is much less likely to be responded to with a punch when boys are playing than when they are fighting. In a statistical model, this is easily represented as an interaction effect between the previous action type and the operative frame, but this begs the question of how we know which frame is operative at a given moment, for while frames are frequently anchored to particular locations, frames can also switch (Goffman 1974), sometimes with little to mark the transition.

Another aspect of content is topic. Conversation analysts have written on how topic talk is conducted (e.g., Sacks 1995, 1:535–42), but with little attention to the interactional consequences of topics of different sorts. Topics are, in fact, important both for what is said in particular turns and for who is selected to say it (Dovidio et al. 1988), as well as for levels of engagement, feelings of solidarity, levels of post-encounter self-confidence, and subsequent odds of reencountering the same people (Collins 2004). But topics are not always explicitly restated in individual turns, and topic transitions are sometimes gradual rather than sudden and marked (Jefferson 1984b, Schank 1977).

Concurrency

Although it may be true that, predominantly, one person speaks at a time in conversation (Sacks et al. 1974, Dabbs et al. 1987), a face-to-face encounter abounds in concurrent behavior. Some of this is vocal, such as backchanneling, but much of it is nonvocal and includes gestures, facial expressions, self-grooming, changes in gesture and bodily orientation, and movement related to manual tasks. This refers to behavior on the part of the speaker, but also behavior on the part of the addressee, as well as the behavior of whoever else is present.

Conversation analysis has arguably benefited by largely steering clear of such matters, even if the original explanation was simply that they were not captured in audio recordings (Sacks 1984). Kendon (1990, chapters 4, 7) illustrates the difficulties of trying to describe and make sense of concurrent nonverbal behavior, an effort that can become mired in descriptions with little analytical payoff. By starting with the assumption that many of the practices people use in conversation are adaptations to its serial nature, and the related assumption that what gets said now is likely to be constrained by what was said last, conversation analysis was able, from the start, to ignore much of what happens when people talk and thereby focus its energies on those practices and how they enable the gradual realization of courses of action.

Conversation analysts have not, however, entirely shrunk from the study of nonverbal behavior (e.g., Schegloff 1998, Raymond & Lerner 2007), and while their observations can come across as somewhat impressionistic (e.g., Schegloff 1984), there is considerable evidence that such concurrent behavior is consequential for what gets said (e.g., Goodwin 1980, 1989). But the way forward in terms of modeling is not obvious. If we are simply concerned with dyadic interaction and a single binary description of each participant (e.g., talking or not talking, looking at the other person or not),
time series or event history methods can be used (Gardner & Griffin 1989), with each participant modeled as responding to the immediate past. Matters become more complicated when more than two people are present, when a person can behave concurrently with themselves (for instance, when someone gestures while talking), and when the behaviors are not binary but multinomial [imagine a taxonomy of facial expressions (Ekman & Friesen 1975) or gestures].

CONCLUSION

In describing his brand of (mainly qualitative) interactional analysis, Kendon (1990) writes, “Statistical analyses are not employed because the underlying assumption is that participants in interaction are not the victims of external pushes and pulls of motivation and circumstance. They are, rather, creatures capable of employing strategies of action and follow shared principles to do so” (p. 47). In fact, people are, to some extent, victims of conversation, sometimes pulled away from saying things they wanted to say, or from saying them in the most auspicious sequential environments (Gibson 2000), and sometimes pushed to say things that they had no interest in saying (Sacks 1995, 1:541–42). That does not, however, mean that there are no strategies of action and much less that there are no shared principles, but rather that such strategies have to be responsive to constraints emanating from the sequential environment and the shared principles people apply when evaluating action alternatives in light of it.

The point of quantification is not to reduce people to the status of billiard balls, but to discern patterns not evident in individual episodes of talk, including those stemming from things that people arrive to conversation with, like formal roles, sex, and network positions. The challenge is to study such effects in a way that is faithful to how conversation is progressively, incrementally produced, in line with constraints left behind by recent talk. It is sometimes observed that standard statistical models make unrealistic assumptions about the phenomena they are used to analyze (e.g., Sørensen 1998). This is certainly no less true in the study of microinteraction than in the study of status attainment or social influence. More sophisticated models may prove adequate yet, but only once we come to terms with certain facts: that the set of possible actions is always changing in conversation, that actions and actors (speakers) are selected simultaneously, and that an opening seized by one person is an opening denied to others.

DISCLOSURE STATEMENT

The author is not aware of any biases that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

For comments on earlier drafts of this paper, I am grateful to Ivan Chase, Ann Mische, Geoffrey Raymond, Mary Rose, Regina Smardon, Melissa Wilde, and an anonymous ARS reviewer. All remaining deficiencies of the paper are entirely my responsibility.

LITERATURE CITED


Lerner GH. 1996b. On the 'semi-permeable' character of grammatical units in conversation: conditional entry into the turn space of another speaker. See Ochs et al. 1996, pp. 238–76


Schegloff EA. 1996b. Turn organization: one intersection of grammar and interaction. See Ochs et al. 1996, pp. 52–133


Strodteck FL, Mann RD. 1956. Sex role differentiation in status deliberations. *Sociometry* 19:3–11
## Contents

### Prefatory Chapters

Reproductive Biology, Technology, and Gender Inequality:
   An Autobiographical Essay
   
   **Joan N. Huber** ......................................................... 1

From Mead to a Structural Symbolic Interactionism and Beyond
   
   **Sheldon Stryker** .................................................. 15

### Theory and Methods

Methodological Memes and Mores: Toward a Sociology
   of Social Research
   
   **Erin Leahey** .................................................... 33

### Social Processes

After Secularization?
   
   **Philip S. Gorski and Ateş Altınordu** .......................... 55

### Institutions and Culture

Religion and Science: Beyond the Epistemological Conflict Narrative
   
   **John H. Evans and Michael S. Evans** ......................... 87

Black/White Differences in School Performance: The Oppositional
   Culture Explanation
   
   **Douglas B. Downey** ............................................. 107

### Formal Organizations

Sieve, Incubator, Temple, Hub: Empirical and Theoretical Advances
   in the Sociology of Higher Education
   
   **Mitchell L. Stevens, Elizabeth A. Armstrong, and Richard Arum** .......................... 127

### Political and Economic Sociology

Citizenship and Immigration: Multiculturalism, Assimilation, and Challenges to the Nation-State
   
   **Irene Bloemraad, Anna Korteweg, and Gökçe Yurdakul** .......................... 153
Differentiation and Stratification
The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit, and Consumer Markets
Devah Pager and Hana Shepherd ................................................................. 181
The Second Generation in Western Europe:
Education, Unemployment, and Occupational Attainment
Anthony F. Heath, Catherine Rotbon, and Elina Kilpi .................................. 211
Broken Down by Race and Gender? Sociological Explanations of New Sources of Earnings Inequality
Kevin T. Leicht ............................................................................................. 237
Family Structure and the Reproduction of Inequalities
Sara McLanahan and Christine Percheski ....................................................... 257
Unconscious Racism: A Concept in Pursuit of a Measure
Hart Blanton and James Jaccard .................................................................... 277

Individual and Society
Horizontal Stratification in Postsecondary Education:
Forms, Explanations, and Implications
Theodore P. Gerber and Sin Yi Cheung .......................................................... 299
Gender Inequalities in Education
Claudia Buchmann, Thomas A. DiPrete, and Anne McDaniel ......................... 319
Access to Civil Justice and Race, Class, and Gender Inequality
Rebecca L. Sandefur ....................................................................................... 339
How the Outside Gets In: Modeling Conversational Permeation
David R. Gibson ............................................................................................. 359
Testing and Social Stratification in American Education
Eric Grodsky, John Robert Warren, and Erika Felts ......................................... 385

Policy
Social Networks and Health
Kirsten P. Smith and Nicholas A. Christakis ..................................................... 405

Sociology and World Regions
Gender in African Population Research: The Fertility/Reproductive Health Example
F. Nii-Amoo Dodoo and Ashley E. Frost ......................................................... 431
Regional Institutions and Social Development in Southern Africa
Matthew McKeever ....................................................................................... 453
Conditional Cash Transfers as Social Policy in Latin America:
An Assessment of their Contributions and Limitations [Translation]
Enrique Valencia Lomelí ................................................................. 475

Las Transferencias Monetarias Condicionadas como Política Social en
América Latina. Un Balance: Aportes, Límites y Debates
[Original, available online at http://www.annualreviews.org/go/EValenciaLomeli]
Enrique Valencia Lomelí ................................................................. 499

Indexes

Cumulative Index of Contributing Authors, Volumes 25–34 .................. 525
Cumulative Index of Chapter Titles, Volumes 25–34 .......................... 529

Errata

An online log of corrections to Annual Review of Sociology articles may be found at
http://soc.annualreviews.org/errata.shtml