Journal of Personality and Social Psychology
2003, Vol. 84, No. 4, 857–870

0022-3514/03/$12.00 DOI: 10.1037/0022-3514.84.4.857

The Sounds of Social Life: A Psychometric Analysis of Students' Daily Social Environments and Natural Conversations

Matthias R. Mehl and James W. Pennebaker
University of Texas at Austin

The natural conversations and social environments of 52 undergraduates were tracked across two 2-day periods separated by 4 weeks using a computerized tape recorder (the Electronically Activated Recorder [EAR]). The EAR was programmed to record 30-s snippets of ambient sounds approximately every 12 min during participants' waking hours. Students' social environments and use of language in their natural conversations were mapped in terms of base rates and temporal stability. The degree of cross-context consistency and between-speaker synchrony in language use was assessed. Students' social worlds as well as their everyday language were highly consistent across time and context. The study sheds light on a methodological blind spot—the sampling of naturalistic social information from an unobtrusive observer's perspective.

Although most social psychologists would agree that research should ultimately lead to insight into real-life phenomena, there is a tradition of bias toward decontextualized laboratory methods. Recently, serious concerns have been raised that psychology has lost contact with naturally occurring social life and that the discipline would benefit from a course correction toward a more context- and culture-sensitive psychology (Funder, 2001; Hogan, 1998; Rozin, 2001).

A central tenet in the field is that social contexts influence behavior. However, it is also clear that in real life there is no random assignment to social contexts. Humans are active agents in selecting and shaping their environments, social situations, and interactions (Allport, 1937; Buss, 1987; Swann, 1987)—presumably to maximize person–situation fit. According to Ickes, Snyder, and Garcia (1997),

> Once individuals are in their chosen situation, their *words and actions* [italics added] are genuine reflections of their personalities, and the fact that they display these behaviors in settings they have specifically chosen ensures a substantial degree of consistency in their behavior. (p. 166)

Substantial progress has been made in mapping everyday experiences, and from this it has become increasingly clear that a coherent personality does emerge in daily life—at least in the minds of the research participants (Csikszentmihalyi & Larson, 1984; Diener & Larson, 1984; Epstein, 1979; Fleeson, 2001; Moskowitz, 1994). In other words, when one looks at people's lives from "their," that is, the agent's, perspective, what one sees is a substantial degree of temporal stability and cross-context consistency. A different—and certainly equally important—question, however, is how social lives appear from a perspective that most people naturally adopt most of the time: The perspective of an unobtrusive observer (Funder & Colvin, 1997; Hogan, 1982). How stable are people's daily lives from this "outside" perspective? Is consistency the exclusive domain of the agent, or is it also accessible to the people around him or her?

In this study, we adopted this largely neglected perspective by tracking people's daily lives with the Electronically Activated Recorder (EAR; Mehl, Pennebaker, Crow, Dabbs, & Price, 2001). The EAR is an event-sampling tool specifically designed for the naturalistic assessment of acoustic behavioral traces. Our goal for this project was to track people's social lives from an unobtrusive observer's point of view and to identify the degree of stability across time and situations. Conceptually, we adopt the notion that social life can be understood in terms of two major sources: people's moment-to-moment environments and their natural conversation with people around them. We further make an argument that the way language is used in natural conversation carries rich and valuable information about a person's social life. First, however, we start out by pointing to a methodological blind spot in naturalistic person–environment interactions.

Methodological Issues in Assessing Naturalistic Person–Environment Interactions

Because of their incomparable efficiency, retrospective self-reports have dominated the assessment of everyday life. However, it has also become clear that an overreliance on people's accounts of themselves can come at rather high costs. Asking participants to accurately recall settings, activities, interactions, and experiences over extended periods of time is a task that is largely incompatible with human autobiographic memory (Schwarz & Sudman, 1993; Stone et al., 2000; Tourangeau, Rips, & Rasinski, 2000).

Preparation of this article was aided by National Institutes of Health Grant MH52391 and by a scholarship from the German National Scholarship Foundation. We thank David Buss, Kenneth Craik, Samuel Gosling, Anna Graybeal, and Sean Massey for their helpful comments on previous versions of this article. We are indebted to Geetha Desikan, Kim Fisher, Beth Henary, Allison McClure, John McFarland, Kristy Orr, Nathalie Shook, Monica Sidarous, Mary-Beth Sylvester, and Sara Webb for their help in collecting the data and transcribing the audiotapes.

Correspondence concerning this article should be addressed to either Matthias R. Mehl or James W. Pennebaker, Department of Psychology, University of Texas, Austin, Texas 78712. E-mail: mehl@psy.utexas.edu or pennebaker@psy.utexas.edu

Over the last 15 years, experience sampling (ES) or momentary assessment techniques have evolved as valuable alternatives and gained substantial power and popularity (Csikszentmihalyi & Larson, 1987; Hormuth, 1986; Stone & Shiffman, 1994). People are prompted several times a day and asked to provide instant reports on their momentary experiences, activities, and environments. There is general agreement that minimizing the time lag between the occurrence and the registration of events in question successfully bypasses memory-related problems of retrospective self-reports. Consequently, momentary assessments can be considered relatively objective accounts of a person's daily experiences.

By definition, however, self-reports—both retrospective and momentary—are subject to two important constraints: First, any recall of events is necessarily a subjective construal, a reflection of how the person interpreted the event. This is apparent when the researcher seeks to demonstrate coherence on the level of the psychological meaning of an act, an event, or an interaction. However, theoretically as well as pragmatically it is equally important to ask to what degree a person's everyday life is stable and consistent from a neutral observer's perspective (Funder & Sneed, 1993; Gosling, John, Craik, & Robins, 1998; Hogan, 1982). Are there idiosyncratic traces in people's daily social worlds that are detectable from an outside point of view, or does coherence mainly emerge on the subjective interpretational level?

A second conceptual constraint of any self-report assessment is that people can only recall what they are aware of. People are constantly surrounded by an infinite amount of potential information. Although some of this potential information is processed by the brain, only a small portion passes sensory filters to enter awareness. Aspects of people's social worlds that fall outside their immediate awareness are inevitably lost.

These considerations have inspired observational approaches in the study of person–environment interactions. Barker and Wright (1951) pioneered such a methodology in their famous case study of Raymond, a 7-year-old boy, who was followed by observers over an entire day. Craik (2000), in what he called *lived-day analyses*, extended Barker and Wright's paradigm by following a person with a video camera and obtaining a comprehensive record of the person's social encounters over an entire day. Naturalistic observation studies yield an immense amount of rich information about a person's life. However, the unique advantage is also the most critical pitfall. The amount of data that is collected has restricted this method to ideographic analysis. Another problem with existing observational techniques is that the act of observation always constitutes a major intrusion into participants' worlds and thus reactively influences target behaviors (Barker & Wright, 1951; Craik, 2000).

Our recently developed tool, the EAR (Mehl et al., 2001), combines aspects of both lived-day analyses and ES approaches. The EAR is a microcassette recorder that tracks ambient sounds in people's environment by recording 30-s snippets every 12.5 min over a period of up to 4 days. Participants wear the device while going about their daily lives. In the recordings, the EAR adopts the unique observer perspective of lived-day studies. In restricting the data collecting to only a representative subset of everyday behavior, it incorporates the relative economic advantage of ES strategies. This combination allows for a nomothetic study of everyday social life while preserving naturalistic data at the level of the raw recording. Also, because the EAR operates imperceptibly, measurement-induced intrusions are minimal.

The Sounds of Social Life: Acoustic Sources of Social Information

What can the EAR tell about a person's social life? Much of the social world is represented acoustically and thus is potentially detectable by the EAR. Two major sources of information are readily available on the recordings and are inherently social in nature: people's daily social environments and their natural conversations.

Conceptually, people's social environments can be dissected into where time is spent, what time is spent doing, and with whom time is spent (Csikszentmihalyi & Larson, 1984). The ambient sounds captured by the EAR allow for the reliable coding of people's locations (e.g., inside an apartment, in a public place) as well as their activities (e.g., surrounded by sounds from TV, a lecturer, from cooking, or computer typing). Finally, EAR recordings can capture important information about people's social interactions: Are they alone or with others? Are they engaged in a conversation? Are they on the phone or talking face to face?

Obviously, the categorization of ambient sounds has its limits and cannot reach the degree of differentiation achieved with activity and environment questionnaires (e.g., Csikszentmihalyi & Larson, 1984; Robinson, 1985). However, EAR data allow the calculation of time-use estimates on the basis of acoustic traces uninfluenced by participants' perceptions. Also, the EAR creates the opportunity to attend to subtle social phenomena, such as personal interaction styles and preferences, that participants typically are not aware of when filling in a questionnaire. We readily acknowledge that the EAR is not necessarily a more objective event-sampling tool. Rather, it provides the researcher with a unique perspective, that of an unobtrusive companion who has access to a source of social information that the participant does not have. Thus, different pictures emerging from the EAR as compared with traditional self-report data call for a study of how the two viewpoints differ rather than for a competition for the most accurate account.

Natural conversations constitute the second source of information documented by the EAR. Considering that a wide array of psychological constructs include predictions about how individuals interact with others, it is surprising that stylistic aspects of our everyday conversations have largely been neglected in the study of person–situation interaction (for some of the few exceptions, see Dunbar, Duncan, & Marriott, 1997; Emler, 2001; Reis & Wheeler, 1991). The idea that language use—or, more specifically, word choice—can reveal psychological information about the speaker is not new (for a broad review, see Pennebaker, Mehl, & Niederhoffer, 2003). As early as 1942, Fillmore Sanford proposed that verbal behavior can serve as a powerful personality marker (Sanford, 1942). Several researchers since then have linked differences in word use to aspects of people's self-reported personality, including the Big Five (Pennebaker & King, 1999), dispositional mood characteristics (Weintraub, 1989), self-esteem (Bosson, Swann, & Pennebaker, 2000), or self-monitoring and Machiavellianism (Ickes, Reidhead, & Patterson, 1986).

Language use also reflects aspects of the social situations that people are in (e.g., P. Brown & Fraser, 1979; Forgas, 1985). In

careful sociological analyses, Erving Goffman (1981) identified language use as an impression-management tool that helps negotiate different aspects of ourselves in different social settings. For Goffman, the words that people speak are often not their own but instead reflect transient situational role constraints rather than stable personal characteristics. For example, the degree of formality of a social setting (e.g., R. Brown & Gilman, 1960), power disparities between the interactants (e.g., Morand, 2000), or the intimacy of a relationship (e.g., Weintraub, 1989) clearly affect features of language use.

Taken together, analyses of language use can provide a window to access a broad spectrum of psychologically relevant information about a person's social world. Methodologically, this approach constitutes an implicit assessment strategy. Word choice—compared with a person's locations and activities—is a subtle source of social information that normally falls beyond a person's awareness and control. In sampling snippets of natural conversation, the EAR thus provides a unique opportunity to study automatic linguistic processes in naturalistic settings.

In sum, the tracking of people's daily life with the EAR captures a wide spectrum of important social information. Such a system helps researchers adopt a perspective that has largely been inaccessible in the study of naturalistic psychological phenomena.

Purpose of the Study

Given the novelty of the method, the current project sought to determine the basic psychometric properties of students' everyday social lives as assessed by the EAR. On two separate occasions, participants wore the EAR continuously for 48 hr on normal weekdays. The two monitoring periods were scheduled 4 weeks apart.[1] Three broad research questions around people's everyday social environments and language use in their natural conversations guided the analyses: (a) How stable are people's social environments over time? (b) How reliable is people's language use over time and across interactants? (c) How stable is language use across different social contexts?

Assessing the Stability of Social Environments

Base rates on how, where, and with whom people spend their days are available from national and international time-budget studies (Robinson, 1977; Szalai, Converse, Feldheim, Scheuch, & Stone, 1972) as well ES studies with adolescents and adults (Csikszentmihalyi & Larson, 1984; Larson, 1990; Reis & Wheeler, 1991). Whereas these estimates are based on retrospective or momentary self-reports and thus represent people's perception of their daily lives, this study sought to establish base rates for people's everyday locations, activities, and conversations derived from unobtrusive observation data.

If individuals actively choose and shape their social worlds, people's everyday social environments should be characterized by a substantial degree of stability across time. Research evolving around the person–situation debate has unambiguously demonstrated the stability of people's everyday behaviors when behavior is aggregated in psychometrically sound ways (e.g., Diener & Larson, 1984; Epstein, 1979; Moskowitz, 1982; Nezlek, 1993; Reis, Lin, Bennett, & Nezlek, 1993). Test–retest correlations typically range somewhere between .50 and .80 depending on the behavior in question and the mode of aggregation. Fleeson (2001) recently demonstrated that when participants rate their momentary experiences in terms of Big Five relevant states, the resulting aggregated distributions show almost perfect stability. Unaffected by large daily within-person variability, the degree to which people rated themselves as acting in social, extraverted, or agreeable ways was highly reliable over time.

However, most of the existing research addresses the stability question from what we have described as the agent's perspective (Hogan, 1982). In the ES studies presented by Fleeson (2001), for example, extraverted states were highly stable, but it is not clear (a) what specific acts different participants considered extraverted or (b) how stable these specific acts would be from the observer's perspective. Thus, Fleeson's results, as with most of the research in this area, address coherence of social behavior on the level of subjective interpretation. With the EAR, we can start to determine the degree of stability in peoples' everyday social environments and social behaviors from the outside perspective, a perspective that people naturally adopt, for example, when they form impressions about others.

Assessing the Stability of People's Natural Language Use

Whereas some research is available on the content of daily conversations (Bischoping, 1993; Dunbar et al., 1997; Emler, 2001; Landis, 1927; Moore, 1922), there is virtually no information available on how people naturally talk in everyday life. Base rates for written language use across different genres are available from Biber (1988) and Pennebaker and King (1999). Gleser, Gottschalk, and John (1959) presented frequency information on a series of linguistic variables that were sampled from 5-min speech in a laboratory setting. This study sought to establish base rates for spoken language use derived from the whole spectrum of people's spontaneous everyday conversations.

Only three studies so far have investigated the temporal stability of language use. Pennebaker and King (1999) have shown that written language use in diaries, class assignments, and even professional journal abstracts is surprisingly reliable over weeks and even years—from the use of pronouns and articles to words reflecting emotional tone. Gleser et al. (1959) found evidence for the stability of spoken language over a very short time period—two successive 2-min intervals taken from a 5-min free speech. Across 21 language categories (e.g., word count, adjectives, substantives, pronouns, feelings) the average correlation was .51. Finally, Schnurr, Rosenberg, Oxman, and Tucker (1986) provided further support for the temporal stability of spoken language use by reporting high within-person rank-order correlations based on 83 linguistic variables over a period of 1 week. It is important to note, however, that this statistical approach can only be consid-

[1] Three writing sessions about either a personal trauma or a neutral control topic were inserted midway between the two monitoring periods. Originally, the EAR was implemented to identify subtle social mediators of the health benefits typically experienced after emotional writing (Pennebaker & Graybeal, 2001). Because people's social lives from the EAR perspective constitute scientifically completely unexplored terrain, we decided to use the data to first lay the psychometric foundations for future EAR studies. Results with respect to the writing intervention will be published elsewhere.

ered indirect evidence of temporal stability, because it is inherently confounded with natural differences in base rates between linguistic variables. This study sought to establish the stability of spoken language over a longer period of time, a period of 4 weeks, and derived from a random sample of a person's everyday spontaneous conversations.

A related question concerning reliability of language use concerns the degree to which the speaker's word use is correlated with the target's language. Communication accommodation theory (Giles, Coupland, & Coupland, 1991) posits that a smooth communication requires that both interactants adapt to each other's communicative behaviors. A recent study analyzed the ways dyads used language in chat room conversations and, in the case of the Watergate tapes, how Richard Nixon interacted with three of his aides. In each of the settings, both communication partners tended to use similar linguistic forms (Niederhoffer & Pennebaker, 2002). How much overlap is there between people's language use in their everyday conversations and the language used by those they are talking to? The current study sought to address this question on a cumulative level by comparing participants' word use with the words sampled from people engaged in conversations with the participants.

Assessing the Stability of Language Use Across Social Context

There is very little doubt that language changes as a function of situation (P. Brown & Fraser, 1979; Forgas, 1985; Pennebaker et al., 2003). Not surprisingly, for example, having students eavesdrop and record snippets of conversations between strangers, Cameron (1969) found that the use of profanity is substantially elevated in leisure as compared with job settings. The EAR allows assessment of how language use changes when people change their social environment. For example, do people talk differently when they are at home as compared with in a public place? Is the language people use in an amusement context different from the language they use at work? Do phone conversations require a different language than personal interactions? Admittedly, our examination of stability of language use across social context is restricted, given the ways the EAR allows us to categorize social contexts.

To summarize, the purpose of this first and largely exploratory EAR study was to determine the basic psychometrics of people's everyday social environments and language use in their natural conversations as assessed from an unobtrusive observer's perspective. For both sources of information, we will present base-rate information and test–retest stability coefficients. For people's natural language use, we will also include an assessment of how much between-speaker consistency exists in linguistic features of everyday conversations. Finally, we will explore linguistic consistency across different social contexts.

Method

Participants

Fifty-four introductory psychology students at the University of Texas at Austin signed up for the study. Two students decided not to participate, one because of the 4-week time commitment, the other because of concerns about the EAR. The remaining 52 undergraduate students (28 women, 24 men; mean age 19.0 years, $SD = 1.3$) participated for course credit. All of them completed the experiment.

EAR System

The version of the EAR used in this study consisted of a microcassette recorder (OPTIMUS Micro-32, Fort Worth, TX), an external microphone (OPTIMUS Omnidirectional Tie Clip Microphone, Fort Worth, TX), and a controller microchip. The microchip was programmed with a 30 s on, 12.5 min off cycle[2] and thus produced roughly five intervals per hour. It was impossible for the participant to sense when the recorder was on or off. Participants carried the EAR in a small shock-protected case either attached to their belt (similar to a cell-phone case) or around the shoulder (like a purse). An external microphone was clipped to the lapel of a jacket or the collar of a shirt. The EAR was switched off overnight. For further details on the development and testing of the device, see Mehl et al. (2001).

Procedure

Participants were asked to wear the EAR twice for 48 hr separated by 4 weeks. The monitoring sessions were scheduled either from Monday morning to Wednesday morning or from Wednesday afternoon to Friday afternoon. For each participant, both monitoring periods were originally arranged on the same schedule. This matching was successful for 37 of the 52 students (71%).

First monitoring period. Participants were run in groups of 2 to 4. On arrival at the lab, students were introduced to the procedure of the study. Considerable effort was taken to provide a thorough explanation of the EAR. It was emphasized that the sampling pattern resulted in only 4% of their day being captured and that the short duration of the recording blocks resulted in only brief snippets rather than complete conversations being captured. All participants were encouraged when they returned their EARs to listen to their recordings and erase any parts they considered objectionable or that they preferred to remain private before the researchers listened to the recordings. On completion of a questionnaire package, participants were handed the EAR equipped with a 90-min microcassette tape. Because of restrictions in tape length, participants had to flip the tape after 24 hr of monitoring. All 52 participants remembered to do so within a tolerance period of ± 2 hr.

When the participants returned to the lab 2 days later, the experimenter offered them the option of listening to their recordings.[3] A 7-item questionnaire obtained feedback on the experiences with the system. On a 5-point rating scale from 1 (*not at all*) to 5 (*a great deal*), participants rated the degree to which they were aware of the EAR, the degree to which wearing the system changed their actual behavior, and the extent to which other people recognized the EAR.

Second monitoring period. The procedure for the second monitoring session was virtually identical to the first one. At the end of the session, participants were thoroughly debriefed about the purpose of the study.

Data Preparation

Participants' social environments were coded from the captured ambient sounds. All recorded conversations were transcribed and submitted to linguistic analyses. The coding and the transcribing were done simultaneously.

---

[2] At the onset of each recording sample, a 5-s period is required for the recording volume to reach normal levels. This delay, experienced while internal recorder components approach their quiescent operating state, leaves about 25 s of useful data.

[3] It is interesting, considering the degree of intimacy of the data, that out of the 52 students, only 1 decided to actually check parts of the recording, and this student did not erase anything.

Judges' coding of social environment. Research assistants listened to the complete recordings (both monitoring periods were coded by the same person). At the end of each interval, they coded the social environment using the Social Environment Coding of Sound Inventory (SECSI). The SECSI is a coding system that comprises the person's current *location* (e.g., in apartment, outdoors, in transit), *activity* (e.g., listening to music, on the computer, eating), and *interaction* (e.g., alone, on the phone, talking to others). A list of categories is contained in Table 1.

In addition to acoustic cues such as the noise of a running engine (in transit), the sound of wind blowing (outdoors), typing noises (computer), or the voice of a lecturer (lecture), judges used context information from previous and consecutive intervals to increase their accuracy. For example, if a person, after being on campus (Interval 1) and riding on a bus (Interval 2), enters an apartment (Interval 3), it is inferred that the student has returned home. The certainty of the judgment is then further enhanced by the information from the subsequent recording periods, in which the person might have switched on the TV or gone to the refrigerator to get something to eat.

Judges' reliabilities calculated from a training tape independently coded by all research assistants yielded a mean Cronbach's alpha of .94. For all categories but reading ($\alpha = .12$) and eating ($\alpha = .64$), consistency coefficients were .70 or greater. More detailed information on the SECSI and the training of the judges was given in Mehl et al. (2001).

Tape transcription and linguistic analysis. In addition to coding the social context, research assistants also transcribed all language captured by the EAR. They received special training in dealing with the challenges posed by transcribing oral language, such as handling of repetitions, filler words, nonfluencies, or slang. For each 30-s recording block, the language samples were identified as coming from either the participant (P) or another live person (O) talking with or in immediate proximity of the person. The transcribed language samples were then sorted by speaker (P, O) and submitted to a linguistic analysis using the word-based language analysis program Linguistic Inquiry and Word Count (LIWC; Pennebaker, Francis, & Booth, 2001).

LIWC operates by comparing all words of a text document to an internal dictionary consisting of more than 2,300 words and word stems, falling into over 70 categories (e.g., first-person singular pronoun, article, preposition, negative emotion, social process, present tense). Each word is assigned to those LIWC categories that apply on the basis of its occurrence in the respective subdictionary. For example, the transcribed word *cried* would fall into four categories: "sadness," "negative emotion," "overall affect," and "past-tense verb" (note that most LIWC categorizes are hierarchically organized). Except for total word count, the LIWC output is listed as the percentage of total words per person per monitoring period. Each participant's raw data resulted in 2 (speaker: P, O) $\times$ 2 (first and second monitoring) linguistic analyses.

Table 1
Mapping Students' Social Environment: Judges' Reliabilities and Base Rates for the Electronically Activated Recorder Ratings of Participants' Daily Interactions, Activities, and Locations Across 4 Days of Monitoring

Category	Judges' reliability	Base rate (%)			
		M	*SD*	Min	Max
Interaction					
Alone	.93	68.6	15.0	37.0	97.6
Talking	.99	27.9	12.8	1.7	56.1
To others	.99	24.2	12.3	1.7	55.4
On the phone	1.00	3.8	3.2	0.0	13.3
Laughing	.86	5.9	3.9	0.0	15.8
Activity					
Music on	.89	13.5	9.7	0.0	41.4
TV on	.95	14.6	12.2	1.1	47.5
Computer	.96	8.5	7.4	0.0	38.5
Reading	.12	11.3	10.2	0.0	34.0
Working	.95	5.1	9.7	0.0	37.0
Eating	.64	2.6	2.2	0.0	11.8
Lecture	.98	11.3	6.5	0.0	29.0
Amusement	.71	14.1	18.0	0.0	71.4
Location					
Apartment	.97	56.2	17.9	11.4	86.5
Outdoors	.90	7.2	5.2	1.0	32.9
In transit	.85	4.1	4.2	0.0	16.0
Restaurant	1.00	2.1	3.7	0.0	24.7
Other public places	.86	27.7	14.4	5.3	70.0

Note. $N = 49$; reliability coefficients are Cronbach's alphas based on six transcribers scoring 88 intervals. Min = minimum; Max = maximum.

Results

Evaluation of the EAR Methodology

EAR feedback. At the end of both monitoring periods, participants completed a seven-item questionnaire on their experiences wearing the EAR. Consistently lower item means during the second monitoring period compared with the first indicated habituation over time. Even after the first 2 days of wearing the EAR, however, participants' ratings of its invasiveness were well below the scale midpoints. Across all 4 days, along the 5-point unipolar scale, where 5 = *a great deal*, students rated that they ($M = 2.67$, $SD = 0.81$) as well as people around them ($M = 2.68$, $SD = 0.84$) were moderately aware of the EAR. They reported feeling only slightly uncomfortable wearing the EAR ($M = 2.05$, $SD = 0.70$) and indicated the degree to which the monitor changed their behavior ($M = 1.43$, $SD = 0.52$) and talking ($M = 1.25$, $SD = 0.44$) as minimal. Altogether, the EAR was not particularly distracting nor did it have a significant impact on the participants' social behaviors.

EAR compliance. Participants showed a high degree of commitment wearing the EAR. All participants returned the EAR for their scheduled appointment; no student forgot to flip the tape after 24 hr of monitoring; even the most private conversations and behaviors were captured and (surprisingly) not erased by participants at the end of the study.

On the basis of a targeted monitoring time of 15 hr per day (assuming 8 hr of sleep and an additional hour of not wearing the EAR before going to bed [30 min] and after getting up [another 30 min]) and 4.8 recording periods per hour (12.5-min cycles), the upper limit of intervals per monitoring period was 144. In our sample of 52 introductory psychology students, the EAR recorded an average of 102 ($SD = 34$) intervals during the first monitoring period. Over the second 2 days of monitoring, this mean increased to 113 ($SD = 33$) time blocks. The lower number for the first monitoring period reflects technical problems (unreliable triggering) experienced in the initial phase of the study. Estimates based on debriefing interviews and documented recording times suggest that across all 4 days of monitoring, on average about 15 intervals per participant were lost because of technical problems (e.g., failures of microchip triggers), and about 20 intervals because of participants not wearing the EAR (mostly during sports, exercise, and personal hygiene). Only 2 participants mentioned having turned off the EAR during intimate conversations with their partners (both for less than an hour).

Analyses of Participants' Social Environments

The judges' codings of the participants' social environments were averaged across all intervals of each 2-day monitoring period. The SECSI variables thus represent percentages of all intervals in which a certain category applied.[4] Three participants provided fewer than 50 intervals on one of their two monitoring periods. Their data were excluded from the analyses because of insufficient sampling. The final sample for the analysis of students' daily social environments comprised 49 participants (22 men, 27 women).

Base rates. Table 1 depicts the descriptive statistics for the SECSI categories aggregated over 4 days. It was surprising to learn that students' active social life is restricted to roughly a third of their waking hours. The average participant spent more than two thirds of their waking time surrounded by nothing but background noises. In only about every fourth interval did the EAR capture a participant talking to another person. Students on average were on the phone in another 3.8% of the intervals. Interestingly, a look at the variability of the measures also revealed striking individual differences. Whereas 1 participant was with others 67% of the time, 2 participants spent virtually the entire 2 days alone. A similar picture emerged for the other interaction categories, "talking to others" and "on the phone."

In somewhat less than a third of the intervals, participants exposed themselves to media such as the TV or radio. Lecture attendance was captured in 11.3% of the cases and—even on weekdays within the semester—14.1% of the sound samples were judged by the raters as capturing activities deemed as amusement (e.g., game parlor, sporting event).

In terms of where participants spent their days, judges located them inside their apartments in more than half of the cases; outdoors in 7.2% of the intervals; in transit 4.1% of the time; inside bars, restaurants, or coffee shops in 2.1% of the intervals; and in other public places, including the campus, arcades, malls, or grocery stores, in 27.7% of the cases.

Four-week stability. To assess the reliability of people's environments, simple correlations were computed between amount of time spent within each SECSI category at Time 1 and the time spent at Time 2. As can be seen in Figure 1, correlations for the social context categories showed that students' social environments were characterized by a high degree of stability over a 4-week period. The average test–retest correlation was $r = .54$ for the interaction categories, $r = .50$ for the daily activities, and $r = .65$ for participants' location variables. All but one correlation (eating) was statistically significant ($p < .05$, one-tailed). Overall, 12 out of the 17 correlations were equal to or greater than .50. Interestingly, inherently "social" variables such as being alone, talking to others, laughing, or amusement were among the most stable across the 1-month interval. It is also noteworthy that students' lecture attendance—despite the matching of the weekdays for the first and second recording period—turned out to be only moderately stable.

Analyses of Natural Language Use

As with the social environment analyses, a minimum number of samplings in any given monitoring period was required to ensure reliability of the linguistic analyses. Participants for whom the EAR captured fewer than 50 words in a 2-day period were excluded. The language samples of 5 participants did not meet this criterion. One of these 5 students was also excluded from the analyses of social environment. Note that the remaining 4 students did provide sufficient recording intervals but spoke to virtually no one during the 2-day period. Three of these 4 "nontalkers" (and also the fifth "nonsampler") were men. The final sample for the linguistic analysis consisted of 47 participants (20 men, 27 women).

From the potential 70+ LIWC variables, the 16 variables identified by Pennebaker and King (1999) as sufficiently reliable were retained for the present study. Three additional LIWC variables were included in the current project because of their relevance to spoken (as opposed to written) language: swear words, nonfluencies, and filler words. Finally, the complete set of personal pronouns (first, second, and third person in singular and plural) was included because references to persons provide important information about people's relative positions in their social networks. The final set of 23 LIWC categories, shown in Table 2, was subjected to the linguistic analyses of participants' daily natural language use.

Base rates. Table 2 depicts participants' language use in everyday conversations. Over a 4-day period, the EAR sampled on average slightly more than 1,000 words per person, ranging from 164 to almost 3,000. Roughly every seventh spoken word was a personal pronoun, with almost half of them references to self (*I, me, my*). Participants used swear words in 5 out of 1,000 words. However, the variability between participants was astonishingly high. Whereas 1 person used swear words at the same rate that the average person used positive emotion words, 11 participants (47%) did no cursing at all. Filler words such as *like, well, you know,* or *I don't know* were sampled at a rate of 1.7%. Because of the nature of everyday conversations, participants used social words and present-tense verbs at a high rate.

Four-week stability. The 4-week-stabilities for the linguistic categories are shown in Figure 2. On the basis of a sample size of $N = 47$, 16 out of 23 correlations pass the threshold of significance. The test–retest correlations for the 12 standard linguistic categories averaged $r = .41$, with especially high correlations for categories unique to spoken language (swear words, nonfluencies, filler words). The only exception among the standard linguistic categories was in the use of second-person pronouns ($r = -.10$). The average retest correlation of the seven LIWC categories indicating psychological processes was $r = .24$ and mainly reflected a rather stable use of positive and negative emotion words over time. In addition, the use of present tense was markedly more stable than the other three relativity categories.

[4] At this point a decision had to be made on how to handle "empty" intervals where, because of a lack of ambient sounds, no judgment of context could be made. In 609 of 10,944 intervals (5.6%), judges were able to decipher neither the participants' locations nor their activities or interactions. Separate analysis with and without these intervals yielded comparable results, with consistently slightly higher base rates when calculated from the reduced set of intervals. Because empty intervals constitute a methodological challenge for the EAR rather than a nonbehavior on behalf of the participant, we considered the results based on the nonempty intervals only more accurate proxies of a person's real behavior.

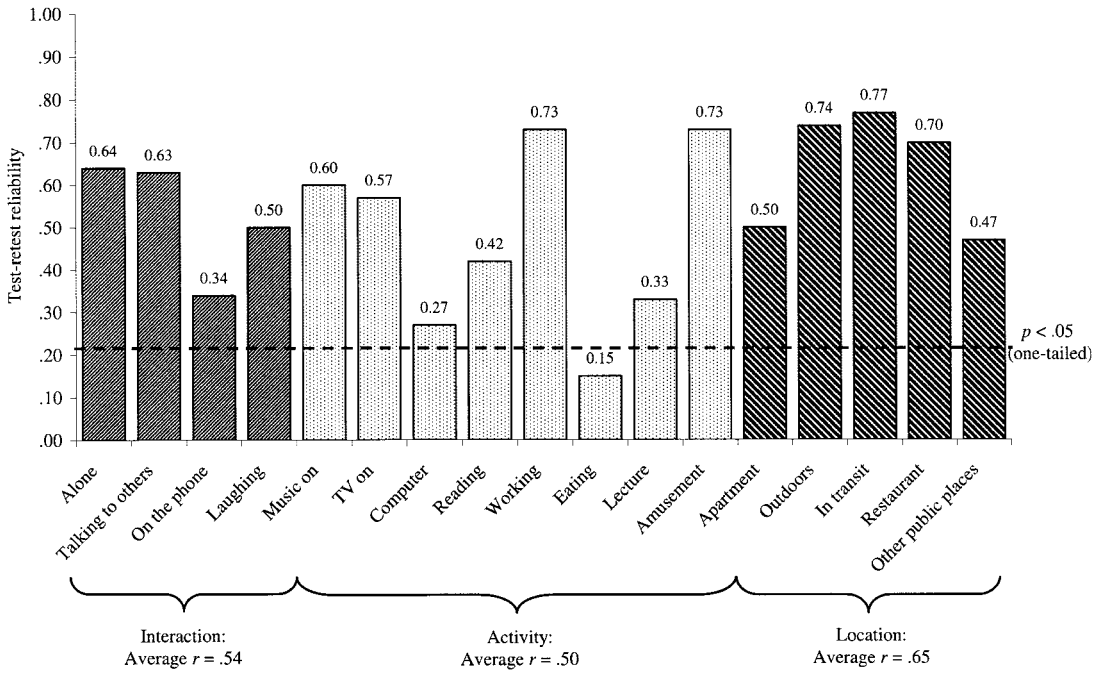

Figure 1. Four-week stability of judges' ratings of social context: Test–retest reliabilities for all 17 Social Environment Coding of Sound Inventory categories. $N = 49$.

Table 2
Students' Language Use in Natural Conversations: Base Rates for Selected LIWC Categories Across 4 Days of Monitoring

LIWC categories	Examples	*M*	*SD*	Min	Max
Standard linguistic dimension					
Sampled raw word count		1,064.4	559.4	146.0	2,943.0
Words of more than 6 letters		8.8	1.7	5.2	13.2
First-person singular pronouns	I, me, my	6.9	1.8	3.2	11.1
First-person plural pronouns	We, us, our	1.0	0.6	0.0	2.6
Total second-person pronouns	You, your	3.7	1.0	1.5	5.7
Total third-person pronouns	She, him, their	2.8	1.1	0.8	5.7
Negations	No, not, never	3.4	1.2	1.6	9.2
Articles	A, an, the	3.9	1.0	1.5	6.2
Prepositions	To, with, above	8.9	1.4	2.8	11.7
Swear words	Damn, bastard	0.5	0.7	0.0	3.4
Nonfluencies	Uh, er	0.4	0.5	0.0	2.4
Filler words	Yaknow, like, Imean	1.7	1.4	0.0	5.7
Psychological processes					
Positive emotions	Happy, good	3.2	1.2	0.5	6.8
Negative emotions	Hate, ugly	1.4	0.8	0.0	4.1
Causation	Because, effect	1.2	0.5	0.3	2.8
Insight	Realize, know	2.2	0.6	0.8	3.6
Discrepancy	Would, should	2.3	0.7	0.8	4.7
Tentative	Perhaps, maybe	2.4	0.8	0.9	5.1
Social processes	Friend, talk	11.1	2.0	5.5	16.8
Relativity					
Past tense	Was, went	4.5	1.5	2.0	7.8
Present tense	Is, go	15.9	2.0	10.9	20.2
Inclusive	With, and	4.9	0.9	3.0	6.8
Exclusive	Except, but	4.1	0.9	1.9	6.0

Note. Base rates (except raw word count) are expressed in percent of total words across 4 days. Filler words or phrases (e.g., *like* or *you know*) were identified by transcribers and converted into uniquely identified single words (e.g., *yaknow, Imean*) that were captured by Linguistic Inquiry and Word Count (LIWC). Min = minimum; Max = maximum.

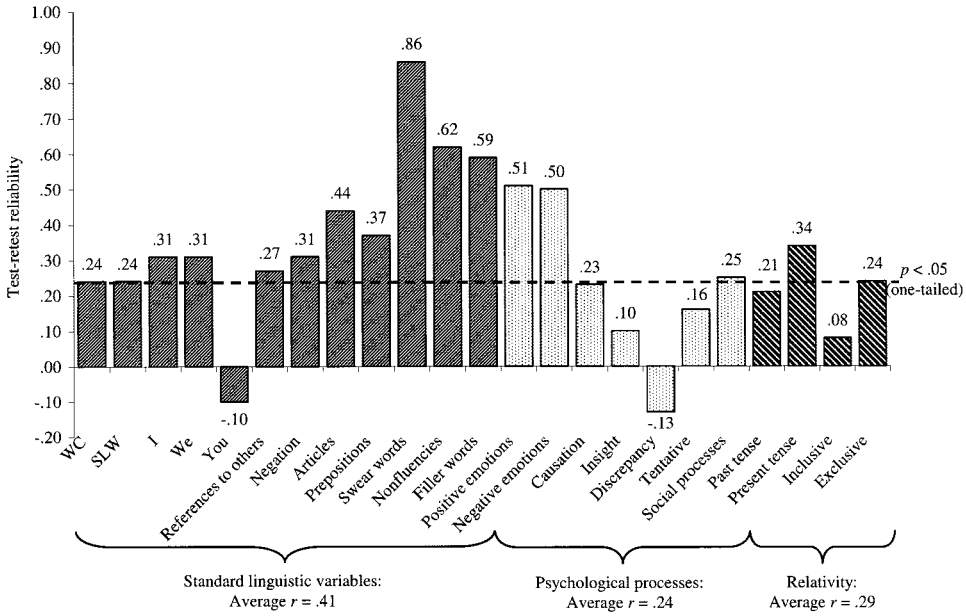

Figure 2. Four-week stability of the linguistic categories: Test–retest reliability for the 23 Linguistic Inquiry and Word Count (Pennebaker et al., 2001) categories. $N = 47$. WC = word count; SLW = words of more than six letters.

Stability between speakers. To test for linguistic synchrony in people's everyday conversations on a general level, we calculated the degree of similarity between the overall language use by the participants and their interaction partners. Recall that in addition to the participants' language (P) we also transcribed when another person (O) was talking with or in immediate proximity of the participant. These O-language samples, irrespective of the actual person they originated from (e.g., friend, romantic partner, stranger) were analyzed cumulatively and resulted in one overall linguistic profile of people talking with the participant.

Figure 3 shows the P–O correlations for the 23 selected LIWC categories on the basis of the entire 4 days of monitoring. With an average correlation of $r = .35$, the standard linguistic LIWC variables show the highest correspondence between speakers. Categories unique to spoken language again range among the categories with the highest correlations (swear words, nonfluencies, filler

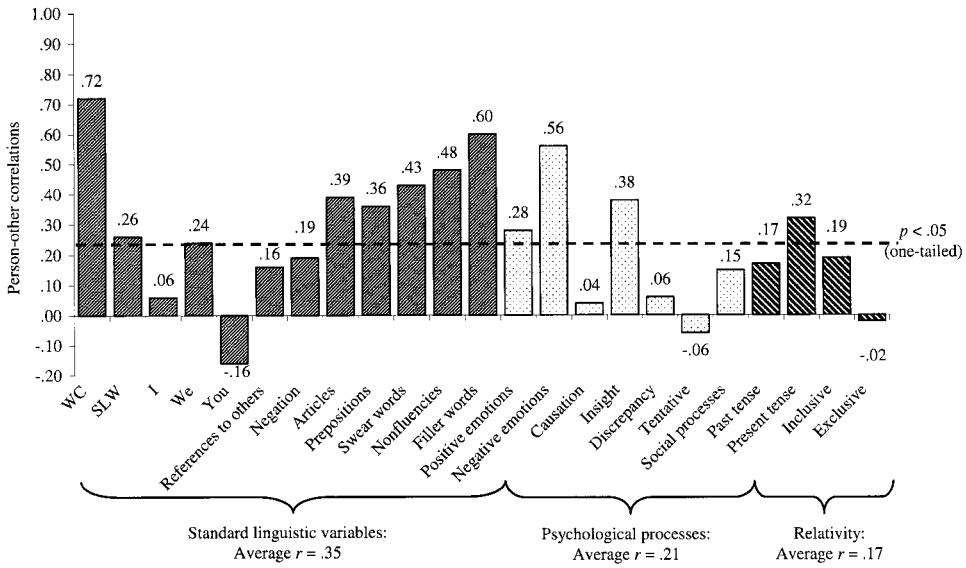

Figure 3. Similarity between the participant's and his or her interaction partners' overall language use: Person–other correlations for the 23 Linguistic Inquiry and Word Count (Pennebaker et al., 2001) categories. $N = 47$. WC = word count; SLW = words of more than six letters.

words). Both overall word counts were highly correlated ($r = .72$). The use of second-person pronouns between speakers was unrelated ($r = -.16$). Interestingly, the use of first-person singular pronouns failed to show P–O correspondence ($r = .06$).

Natural Language Use and Social Context

Students' daily social environments as well as their language use in natural conversation are reliable over time. Is the reliability of language use due (at least to some extent) to situational contexts driving certain kinds of language use? If so, linguistic reliability would emerge as an epiphenomenon of situational reliability. Unfortunately, a strong test of this question was restricted by two factors: our abilities to categorize situations and the comparatively small sample size. The following analyses, then, should be considered preliminary.

For the context analyses of language use, a new data set was created where the language uttered by the participant during each interval was submitted to a separate LIWC analysis.[5] The new interval-based LIWC information (e.g., percentage of emotion or swear words during the 30-s EAR recording) was matched with the interval-based information about students' social context derived from the judges' SECSI ratings. This data file was then aggregated by participants and social contexts, so that averaged LIWC data were available for each participant on the social contexts we sought to compare. The analyses are based on the participants who were included in both the linguistic as well as the social context analysis ($N = 45$).

Language use across location. To what extent do participants' locations constrain their language use? For this, the original SECSI location categories "outdoors" and "in transit" were combined into "outside," and "restaurant" and "other public places" were merged into "in public places." The new categories were then contrasted with the category "apartment." A series of within-subject analyses of variance (ANOVAs) on the 23 LIWC variables revealed significant context effects only for word count, $F(2, 80) = 4.34$, $p < .05$, and first-person plural pronouns, $F(2, 80) = 4.21$, $p < .05$. Participants uttered significantly fewer words in a 30-s period when in public places ($M = 18.5\%$) than when in their apartments ($M = 23.4\%$), $t(40) = 3.28$, $p < .01$, or outside ($M = 22.1\%$), $t(40) = 2.16$, $p < .05$. Also, they used first-person plural pronouns at a substantially higher rate in public places ($M = 1.0\%$) as compared with at home ($M = 0.7\%$), $t(40) = 3.94$, $p < .01$.

Language use across activity. A second question is whether students' language use changed as a function of their activity within a specific location. On the basis of the SECSI information, talking in public places occurred mainly while working or pursuing activities deemed amusements. Other activity categories, such as "on the computer," "reading," "eating," or "attending a lecture," had insufficient talking base rates to be included in the context analysis. Accordingly, within the category "in public places," the categories "working" and "amusement" were contrasted with a residual category, "other activities." Unfortunately, only 13 participants provided data on all three activity categories. On the basis of this very restricted sample, within-subject ANOVAs for the LIWC variables revealed significant mean differences solely in first-person singular pronoun use, $F(2, 24) = 5.16$, $p < .05$. Participants referred to themselves more than twice as much in amusement contexts ($M = 10.2\%$) than while working ($M = 3.8\%$); $t(12) = 3.63$, $p < .01$ ($M_{other} = 6.2\%$).[6]

Language use across mode of interaction. Finally, we were interested in the extent to which the mode of interaction, talking to a person directly versus on the phone, determined language use. To address this question, language captured by the EAR while participants were on the phone inside their apartments was contrasted with transcripts of real conversations that were captured at home. Note that we wanted to look at how language changes as a function of social context within a given location. The analyses revealed that language use was independent of the actual mode of interaction. Statistically significant differences only occurred in 4 out of the 23 LIWC categories. On average, in phone as compared with real conversations, participants uttered more words per interval ($M = 30.1\%$ vs. $M = 21.5\%$), $t(39) = 3.70$, $p < .01$; used nonfluencies at a higher rate ($M = 1.0\%$ vs. $M = 0.4\%$), $t(39) = 2.11$, $p < .05$; used fewer first-person plural pronouns ($M = 0.5\%$ vs. $M = 0.8\%$) , $t(39) = -2.03$, $p < .05$; and used fewer inclusive words ($M = 3.4\%$ vs. $M = 4.2\%$), $t(39) = -2.02$, $p < .05$.

Summary. In the cross-context analyses, language use emerged as stable across locations, activities, and modes of interaction. Contrasting participants' word choice at home, outside, and in public places; their language in amusement versus work contexts; and talking styles in personal conversations versus phone calls at home revealed only a very limited number of significant effects. The results, however, should be interpreted cautiously for two reasons: First, the contrasts were limited by the restricted context information available from the SECSI. Second, they were based on a restricted sample because not all participants provided language data across all contexts.

Other Relevant Data

Although not the focus of this article, the data set allowed us to perform exploratory analyses on gender differences in students' daily social environment and language use in natural conversations. Relatively few differences in preferences for social settings and activities emerged. Over a period of 4 days, female as compared with male participants spent significantly more time on the phone ($M_{female} = 4.8\%$ vs. $M_{male} = 2.5\%$, $p < .05$) and less time on the computer ($M_{female} = 6.1\%$ vs. $M_{male} = 11.4\%$, $p < .05$). All other SECSI categories yielded statistically equal mean base rates. Compared with the relatively few differences in preferences for social settings and activities, the language used by male and female participants in their everyday conversations differed in a number of ways. Consistent with previous research (Pennebaker & King, 1999), men used significantly more big words (words more than six letters long; $M_{male} = 9.4\%$ vs. $M_{female} = 8.3\%$, $p < .05$), more articles ($M_{male} = 4.4\%$ vs. $M_{female} = 3.5\%$, $p < .01$), fewer first-person singular pronouns ($M_{male} = 6.2\%$ vs. $M_{female} = 7.5\%$, $p < .01$), and fewer discrepancy words ($M_{male} = 2.0\%$ vs. M_{female}

[5] Following a strategy adopted by Niederhoffer and Pennebaker (2002), we restricted the interval-based LIWC analysis to utterances of at least three words. In 214 out of 2,449 intervals (8.7%), participants were saying fewer than three words. These intervals were excluded.

[6] The results of the post hoc tests are based on the 13 participants who were included in the overall ANOVA. Analyses using pairwise exclusion yielded similar results.

= 2.5%, $p < .05$) than women. Also, men's transcripts contained fourfold the amount of swear words (M_{male} = 0.9% vs. M_{female} = 0.2%, $p < .01$), considerably fewer filler words (M_{male} = 1.3% vs. M_{female} = 2.1%, $p < .05$), and fewer references to positive emotions (M_{male} = 2.8% vs. M_{female} = 3.5%, $p < .05$).

Somewhat surprisingly, male participants used significantly more negative emotion words in their daily language (M_{male} = 1.9% vs. M_{female} = 1.1%, $p < .01$). More detailed analyses revealed that this effect was exclusively produced by men using an almost threefold amount of anger words (M_{male} = 1.1% vs. M_{female} = 0.4%; $p < .01$). No difference emerged in the use of anxiety- and sadness-related words. Also, contrary to the widely held notion of women's higher social relatedness, women's everyday language did not contain more references to social processes or more first-person plural pronouns.

Discussion

The purpose of this study was to determine the basic psychometric properties of people's social lives as measured by their daily social environments and their language use in their natural conversations. Using EAR methodology, we adopted the unique perspective of an unobtrusive observer. The results indicate that students' social lives show a remarkable degree of stability even from this outsider perspective. They support the idea that coherence in personality is not merely an experiential phenomenon in people's minds but is also driven by consistency in objective—in the sense of traceable—aspects of people's everyday social lives.

Students' Daily Social Lives From an Acoustic Perspective

EAR base rates for students' social lives. One of the goals of this study was to determine base rates for people's social environments and natural conversations from the perspective of an unobtrusive observer. By and large, the estimates that emerge from this study map onto base rates obtained from time-budget studies (Robinson, 1977, 1985; Szalai et al., 1972)—predominantly based on retrospective self-reports—and ES studies (Csikszentmihalyi & Larson, 1984; Larson, 1990). This is particularly true for the SECSI location and activity categories. Here, substantial differences only emerged for watching TV (EAR 14.6% vs. ES 7.2%) and listening to music (EAR 13.5% vs. ES 1.4%) and are likely driven by interpretational differences in what constitutes watching TV or listening to music. It is important to note that the SECSI location and activity categories comprise highly observable, that is, acoustically detectable, behaviors and events. Observability has been shown to facilitate self–other agreement (Funder & Colvin, 1997; Gosling et al., 1998).

The picture looks different for the SECSI interaction categories. Both time-budget and ES studies have found that the time adults spend alone (and awake) falls somewhere between 4 hr and 6 hr per day, which is equivalent to 25%–38% of their time awake, whereas the EAR yielded an estimate of 68.6%. ES studies have pinpointed students' amount of daily conversations at around 6 hr—about 38% of the day (Reis & Wheeler, 1991)—whereas the EAR data suggest 27.9%. Interpretational differences, such as what it means to be alone or engaged in a conversation, again emerge as a potential explanation. The EAR, for example, identifies a person studying at a coffee shop as "alone" because there, the person is "a 'single,' a party of one, a person not in a *with*, a person unaccompanied 'socially' by others in some public undertaking" (Goffman, 1981, p. 79), whereas the participants themselves most likely would report being with—in the sense of "surrounded by"—people (e.g., Larson, 1990).

Another possibility is that the timing of the sampling explains the differences in time estimates. That is, the midweek-only sampling in this study versus the weekend-also sampling in typical diary studies (e.g., Fleeson, 2001; Larson, 1990; Reis et al., 1993) could have resulted in differential patterns of social interactions. Surprisingly, although certain qualitative aspects of people's social interactions have been shown to change over the weekend (K. W. Brown & Moskowitz, 1998; Gable & Reis, 1999), no empirical data have yet been published on the actual rates of interactions. An internal analysis of a recently completed data set where 15 people wore the EAR continuously for 10 days revealed no significant differences between the weekday average or any midweek 2-day period and the weekend for "time spent alone" and "time spent in conversation." There was, however, a trend for relatively more dyadic and fewer group interactions over the weekend (Mehl & Pennebaker, in press). Thus, there has been little evidence to date that people change their overall frequency of social interactions over the weekend. Still, the influence of sampling decisions on the time estimates cannot be ruled out. People's weekend social interactions clearly are an interesting topic for future EAR as well as ES research.

Over and above interpretational ambiguities and potential sampling effects, however, it is interesting that the social interaction base rates diverge in a way that makes people's everyday social lives appear more social from the agent's than the observer's point of view. The question then arises whether this is a random effect or due to characteristics inherent to social interactions. Social interactions are the landmarks of social life and stand out as punctuated figures on a rather homogeneous ground of daily activities. Craik (2000) reported an interesting phenomenon in his lived-day analyses: Judges experience a strong impulse to press the fast-forward button when watching nonsocial periods of the recordings of a person's day. This anchoring of people's perceptions of everyday life around their social interactions then could potentially make social episodes the target of well-known biases in human information processing (Schwarz & Sudman, 1993; Stone et al., 2000; Tourangeau et al., 2000). Although obtrusive aspects, such as with whom and about what one had a conversation, probably pose no problem for accurate recall, more subtle aspects, such as the duration of daily conversations, run the risk of being misrepresented and—because of their high salience and self-relevance—most likely overestimated.

This underscores our basic conceptual argument for the EAR: It is not a generally more objective, in the sense of accurate, event-sampling tool. Rather, it is a research tool that allows us to obtain insight into people's social lives from a unique (i.e., the unobtrusive observer's) point of view. This perspective can yield information similar to or quite different from the person's insider perspective, depending on the event or behavior under investigation. Divergence is likely for events or aspects of events that do not naturally capture a person's attention (as, e.g., in subtle, low-frequency, or counterhabitual behavior). A study collecting both EAR and ES data on the same participant

could test this idea systematically and would be extremely valuable for developing a better understanding of the psychological processes involved in self- and other assessment (Funder & Colvin, 1997).

Individual differences in students' social lives. The degree of individual variability in students' everyday social lives was striking. The 52 participants differed tremendously in their preferences for social settings, activities, and interactions. Some participants spoke to virtually no one during the entire 4 days of monitoring; others were engaged in conversations the majority of their waking hours. The ways people used language were similarly variable. Some swore quite a bit, many not at all. In fact, 1 participant used profanity more often (3.4%) than another person made self-references (3.2%). The same applied for emotional words, cognitive words, filler words and most of the other linguistic markers. Obviously, some of the base rates for peoples' social environment and language use, such as the time spent at home or the use of swear words, would vary with the target population. Future research needs to address this issue by identifying moderators such as a person's socioeconomic status, age, and cultural or subcultural background. The most immediate next step, however, is to lend psychological meaning to the individual differences in people's social lives. In a follow-up project, we are currently looking for personality correlates of people's social environment selection and natural word choice. Do extraverts go out more frequently than introverts? Do depressed people socially withdraw by staying at home? Do neurotics whine and complain in their daily conversations? Do narcissists show a linguistic self-focus? This will ultimately enrich personality psychology by providing contextualized personality correlates derived from naturalistic observation data (Funder, 2001; Hogan, 1998).

As an aside, it is interesting that in this study gender accounted for so little variance in students' daily social environments. Although women spent more time on the phone and men were on the computer more often, all other aspects of daily locations, activities, and interactions yielded comparable estimates. Female and male participants, however, differed substantially in their everyday language use. Overall, the differences converge with findings for written language use from this and other labs (Lakoff, 1975; Mulac, Bradac, & Gibbons, 2001; Pennebaker & King, 1999)—with two interesting exceptions. This study did not find support for the notion that women in general refer to emotions more often (for a review, see Mulac et al., 2001). Although women used more positive emotion words in their everyday language, men had a higher prevalence for angry utterances. Also, contrary to the widespread idea that women's language is more socially engaging (Maccoby, 1990; Tannen, 1990), and also contrary to findings concerning written narratives from our lab (Pennebaker & King, 1999), female participants did not make more references to social processes.

Clearly, research on people's everyday spontaneous language use is in its infancy. Future research will have to adopt a more differentiated perspective and systematically distinguish between same-sex and opposite-sex interactions (e.g., Thomson, Murachver, & Green, 2001) as well as address basic pragmatic and functional differences between written and spoken language use (e.g., Biber, 1988; Pennebaker et al., 2003).

Stability of Students' Social Lives

Temporal stability. Analyses of the various SECSI categories revealed a degree of stability in students' daily social environments over a period of 4 weeks that is comparable to the stability found in self-report and observational research (Diener & Larson, 1984; Epstein, 1979; Fleeson, 2001; Moskowitz, 1982; Nezlek, 1993; Reis et al., 1993). The rather low stability for lecture attendance was surprising considering the matching of weekdays for both monitoring periods and the high intercoder reliability. Overall, however, the data clearly revealed that people's everyday lives are not only coherent from the agent's perspective but also show a high degree of consistency from an outsider's perspective.

The average stabilities for the linguistic categories were lower than those for the social environment categories. Considering that the EAR sampled language across a wide spectrum of conversations, with some being very private and some not more than a casual "How are you?" their magnitude was surprising. Moderate to high correlations were found between Time 1 and Time 2 for the use of articles, prepositions, first-person pronouns, and present-tense verbs. The use of positive and negative emotion words was equally reliable. The variety of topics of people's daily conversations suggests that this stability reflects a stability of linguistic style more than linguistic content. The fact that the stability coefficients for the use of swear words, nonfluencies, and filler words exceeded the ones for any other LIWC category shows that spoken language provides valuable stylistic information over and beyond written language (Pennebaker & King, 1999).

What are the implications of the fact that people's social lives are stable from an observer's point of view? It is easy to see that apparently unrelated behaviors such as being at a coffee shop (reading) and typing on the computer (writing a paper) can be coherently interpreted as "studying" on a subjective level. This study, however, shows that over and beyond this "higher order" experiential consistency, there is relative stability in people's social events such as where, with what, and with whom people spend their days. Demonstrating this objective stability means establishing reliability for one of the core criteria of naturalistic personality research: people's everyday behaviors. Real-life behaviors can only be predicted from people's dispositions if these real-life behaviors follow sound psychometric properties—not only in the agent's eyes but also in the eye (or EAR) of a neutral observer.

Demonstrating temporal stability of people's everyday social lives from the observer's point of view is also conceptually important for research on impression formation. Naturally, people form impressions by seeing other people act in their natural environments. Self-reported consistency has been shown to reliably facilitate self–other agreement in personality assessment (Funder & Colvin, 1997). Once consistency in people's objective everyday behaviors and social environments has been established, one can start to determine which cues laypeople use in forming impressions about others. People's everyday social environments and their language use in spontaneous conversations certainly are good candidates for being social cues that people naturally draw on when figuring out who a person is (e.g., Gifford & Hine, 1994; Gosling, Ko, Mannarelli, & Morris, 2002). Related to this idea, it would be interesting for future research to identify what kind of

social life and natural language use make people be seen as likeable by others.[7]

Between-speaker stability. Transcribing utterances by both the participants and their interaction partners made it possible to test between-speaker stability in language on a broad level. Communication accommodation theory (Giles et al., 1991) identifies convergence and divergence as strategies that humans adopt to manage conversations. Our analyses revealed varying degrees of overlap in word choice between speakers. Swear words, nonfluencies, and filler-word categories showed a higher degree of convergence than most other variables. First-person singular and second-person pronouns failed to show convergence, most likely reflecting that they are typically used complementarily in dyadic conversations. The total word counts for both sets of transcripts were highly correlated (.72). Our findings confirm on a cumulative level what Niederhoffer and Pennebaker (2002) found on a turn-by-turn level: Natural conversations are characterized by a substantial level of linguistic synchrony. Unfortunately, this study did not allow us to identify sources of this synchrony. Do people prefer to interact with linguistically similar others? Do people subtly impose their linguistic style onto their interaction partners? Or is synchrony simply a reflection of the conversational context? By whatever source synchrony may finally be driven, it certainly fosters temporal stability in people's linguistic styles.

Cross-context stability. Finally, the EAR data allowed us to look at how participants' language varied across social contexts. Clearly, language use is constrained by situational factors (Forgas, 1985; Pennebaker et al., 2003). Our analyses of word choice, however, also paint a picture of relative stability across context. The comparisons of language use in public places versus at home, in amusement versus work contexts, and on the phone versus in direct personal interactions did not yield significant linguistic differences effects over and beyond what can be expected by chance. These results in combination with the stability of language use across time provide strong support for the idea that people express themselves in idiosyncratic ways using distinctive linguistic styles. Note, however, that the analyses are exploratory in nature for two reasons: (a) The selection of contexts was seriously constrained by what could be coded from ambient sounds, and (b) because not every student provided language samples from all contexts, some contrasts were based on a small number of observations.

Ethical and Legal Issues in Conducting EAR Research

Clearly, recording snippets of people's social lives raises some critical ethical and legal questions. As mentioned in the Method section here and in more detail elsewhere (Mehl et al., 2001), this study implemented several safeguards to protect participants' privacy and ensure maximum data confidentiality. However, the most serious legal and ethical concerns revolve not around the participants themselves, but rather potential "secondary" participants, bystanders who are not directly involved in the study but whose voices and behaviors are captured by the EAR. Although for future studies an optimized EAR system is available that maximizes sound emitted by the participants while suppressing information about people around them (by means of low sensitivity recording with a modified unidirectional microphone), this study did capture and use information from other people.

In the United States, there are very few restrictions about recording people's utterances in public places. The situation concerning the recording of private conversations is far more ambiguous—both legally and ethically. In most states, including Texas, where the current study was conducted, recordings can be made legally if at least one of the interactants has knowledge of the recording device. A small number of states (e.g., Maryland) only allow recordings if all interactants have knowledge of the recording. Even in the most legally restrictive states, however, unauthorized recordings are only a problem if the participants are identifiable.

In the present study, participants did wear the EAR in their own or at friends' houses, and other people were recorded in private settings. However, participants were asked to incorporate the EAR into their daily lives as naturally as possible. This means they were encouraged to wear the microphone visibly and openly talk about the EAR and discuss it with others in their conversations. Anecdotal information from the debriefing sessions revealed that participants regularly informed their friends and partners about the EAR.

Irrespective of notification, anonymity of other people's utterances is of paramount importance, because their speech samples are collected without informed consent. In this study, anonymity of the data was maximized by limiting recordings to 30-s snippets. In no case did the sound samples reveal the full, that is, first and last, name of a person. Also, the 30-s intervals generally captured the last words of an utterance, one or two more complete sentences, and the beginning of another utterance. Substantial context information is necessary to infer the meaning of a sentence within a conversation. It is thus highly unlikely that this study violated privacy rights of a person who was inadvertently recorded.

As further protections, the audiotapes were transcribed either by professional transcribers or research staff who did not have access to the identities of the research participants. In addition, they were trained to omit any potentially identifying information from the transcripts themselves. Recall also that participants were given the option of listening to their tapes and erasing any portions before they turned them over to the experimenters.

Despite all of the steps that were taken to ensure the confidentiality of participants and anonymity of nonparticipants, EAR researchers must be aware of the ethical implications of recording others. Through careful instructions and informed consent procedures with participants, brief recording periods, detailed training of transcribers, and the removal of all identifying information, the present procedure adheres to both ethical and legal standards for research.

Potentials for Future EAR Research

The EAR has great potential for researchers in social and personality psychology. Currently, it is the only available methodology that allows the sampling of behavioral data from an unobtrusive observer's point of view. Although many of the EAR data overlap with what is obtained from retrospective or momentary self-reports, the EAR's unique strength lies in the tracking of subtle acoustically detectable events. Verbal and paraverbal be-

[7] We thank an anonymous reviewer for bringing up this creative idea.

havior in spontaneous conversations clearly emerges as the most immediate arena for the EAR. However, a variety of other real-life phenomena that usually fall outside a person's awareness are also inherently important to personality and social psychologists: Subtle signs of emotions, such as laughter, sighing or whistling; unobtrusive health indicators, such as coughing, sneezing or sniffing; or even ambient sound levels in a person's environment can be detected, quantified and analyzed. Also, the EAR technology poses only pragmatic constraints on how many data points can be obtained from a person. Doubling or even tripling the sampling rate, for example, to more adequately capture low-frequency behavior, constitutes no additional burden on behalf of the participant (ethically though, it probably requires proportionally shorter recording intervals). Finally, as an archival record, the EAR—unlike many methodologies in the social sciences—produces rich records of people's daily lives that are readily available for other researchers and questions that go beyond the investigators' primary interests. It is conceivable that generations of future scientists could have an interest in analyzing the EAR records using their—maybe then very different—interpretive lenses.

EAR research is high-investment research. The data collection and preparation process is time-consuming and labor intensive. The EAR's potential as a tool for unobtrusive naturalistic observation makes it a needed complement to psychologists' long-standing reliance on self-reports. It provides a unique window into a wide variety of real-life social phenomena that ultimately constitute the endpoint of our research endeavors. In its unobtrusiveness and its outsider viewpoint, the EAR calls for the study of those social phenomena that have too long existed in the shadow of the reportable.

References

Allport, G. W. (1937). *Personality: A psychological interpretation.* New York: Holt.

Barker, R. G., & Wright, H. F. (1951). *One boy's day. A specimen record of behavior.* New York: Harper & Row.

Biber, D. (1988). *Variation across speech and writing.* Cambridge, England: Cambridge University Press.

Bischoping, K. (1993). Gender differences in conversational topics, 1922–1990. *Sex Roles, 28,* 1–17.

Bosson, J. K., Swann, W. B., Jr., & Pennebaker, J. W. (2000). Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited? *Journal of Personality and Social Psychology, 79,* 631–643.

Brown, K. W., & Moskowitz, D. S. (1998). Dynamic stability of behavior: The rhythms of our interpersonal lives. *Journal of Personality, 66,* 105–134.

Brown, P., & Fraser, C. (1979). Speech as a marker of situation. In K. R. Scherer & H. Giles (Eds.), *Social markers in speech* (pp. 33–62). Cambridge, England: Cambridge University Press.

Brown, R., & Gilman, A. (1960). The pronouns of power and solidarity. In T. A. Sebeok (Ed.), *Style in language* (pp. 253–276). Cambridge, MA: MIT Press.

Buss, D. M. (1987). Selection, evocation, and manipulation. *Journal of Personality and Social Psychology, 53,* 1214–1221.

Cameron, P. (1969). Frequency and kind of words in various social settings, or what the hell's going on? *Pacific Sociological Review, 12,* 101–104.

Craik, K. H. (2000). The lived day of an individual: A person–environment perspective. In W. B. Walsh, K. H. Craik, & R. H. Price (Eds.), *Person–environment psychology: New directions and perspectives* (pp. 233–266). Mahwah, NJ: Erlbaum.

Csikszentmihalyi, M., & Larson, R. (1984*). Being adolescent: Conflict and growth in the teenage years.* New York: Basic Books.

Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the experience-sampling method. *The Journal of Nervous and Mental Disease, 175,* 526–536.

Diener, E., & Larson, R. J. (1984). Temporal stability and cross-situational consistency of affective, behavioral, and cognitive responses. *Journal of Personality and Social Psychology, 47,* 871–883.

Dunbar, R. I. M., Duncan, N. D. C., & Marriott, A. (1997). Human conversational behavior. *Human Nature, 8,* 231–246.

Emler, N. (2001). Gossiping. In W. P. Robinson & H. Giles (Eds.), *The new handbook of language and social psychology* (pp. 317–338). New York: Wiley.

Epstein, S. (1979). The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology, 37,* 1097–1126.

Fleeson, W. (2001). Toward a structure- and process-integrated view of personality traits as density distribution of states. *Journal of Personality and Social Psychology, 80,* 1011–1027.

Forgas, J. (Ed.). (1985). *Language and social situations.* New York: Springer-Verlag.

Funder, D. C. (2001). Personality. *Annual Review of Psychology, 52,* 197–221.

Funder, D. C., & Colvin, C. R. (1997). Congruence of other's and self-judgments of personality. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 617–656). New York: Academic Press.

Funder, D. C., & Sneed, C. D. (1993). Behavioral manifestations of personality: An ecological approach to judgmental accuracy. *Journal of Personality and Social Psychology, 64,* 479–490.

Gable, S. L., & Reis, H. T. (1999). Now and then, them and us, this and that: Studying relationships across time, partner, context, and person. *Personal Relationships, 6,* 415–432.

Gifford, R., & Hine, D. (1994). The role of verbal behavior in the encoding and decoding of interpersonal dispositions. *Journal of Research in Personality, 28,* 115–132.

Giles, H. G., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of accommodation: Developments in applied sociolinguistics* (pp. 1–68). Cambridge, England: Cambridge University Press.

Gleser, G. C., Gottschalk, L. A., & John, W. (1959). The relationship of sex and intelligence to choice of words: A normative study of verbal behavior. *Journal of Clinical Psychology, 15,* 183–191.

Goffman, E. (1981). *Forms of talk.* Philadelphia: University of Pennsylvania Press.

Gosling, S. D., John, O. P., Craik, K. H., & Robins, R. W. (1998). Do people know how they behave? Self-reported act frequencies compared with on-line codings by observers. *Journal of Personality and Social Psychology, 74,* 1337–1349.

Gosling, S. D., Ko, S. J., Mannarelli, T., & Morris, M. E. (2002). A room with a cue: Personality judgments based on offices and bedrooms. *Journal of Personality and Social Psychology, 82,* 379–398.

Hogan, R. (1982). A socioanalytic theory of personality. *Nebraska Symposium of Motivation, 30,* 55–89.

Hogan, R. (1998). Reinventing personality. *Journal of Social and Clinical Psychology, 17,* 1–10.

Hormuth, S. E. (1986). The sampling of experiences in situ. *Journal of Personality, 54,* 262–293.

Ickes, W., Reidhead, S., & Patterson, M. (1986). Machiavellianism and self-monitoring: As different as "me" and "you." *Social Cognition, 4,* 58–74.

Ickes W., Snyder, M., & Garcia, S. (1997). Personality influences on the choice of situations. In R. Hogan, J. Johnson, & S. Briggs (Eds.),

Handbook of personality psychology (pp. 165–195). New York: Academic Press.

Lakoff, R. T. (1975). *Language and woman's place*. New York: Harper & Row.

Landis, C. (1927). National differences in conversation. *Journal of Abnormal and Social Psychology, 21,* 354–357.

Larson, R. (1990). The solitary side of life: An explanation of the time people spend alone from childhood to old age. *Developmental Review, 10,* 155–183.

Maccoby, E. E. (1990). Gender and relationships: A developmental account. *American Psychologist, 45,* 513–520.

Mehl, M. R., & Pennebaker, J. W. (in press). The social dynamics of a cultural upheaval: Social interactions surrounding September 11, 2001. *Psychological Science.*

Mehl, M. R., Pennebaker, J. W., Crow, M., Dabbs, J., & Price, J. (2001). The Electronically Activated Recorder (EAR): A device for sampling naturalistic daily activities and conversations. *Behavior Research Methods, Instruments, and Computers, 33,* 517–523.

Moore, H. T. (1922). Further data concerning sex differences. *Journal of Abnormal and Social Psychology, 17,* 210–214.

Morand, D. (2000). Language and power: An empirical analysis of linguistic strategies used in superior–subordinate communication. *Journal of Organizational Behavior, 21,* 235–248.

Moskowitz, D. S. (1982). Coherence and cross-situational generality in personality: A new analysis of old problems. *Journal of Personality and Social Psychology, 43,* 754–768.

Moskowitz, D. S. (1994). Cross-situational generality and the interpersonal circumplex. *Journal of Personality and Social Psychology, 66,* 921–933.

Mulac, A., Bradac, J. J., & Gibbons, P. (2001). Empirical support for the gender-as-culture hypothesis: An intercultural analysis of male/female language differences. *Human Communication Research, 27,* 121–152.

Nezlek, J. B. (1993). The stability of social interaction. *Journal of Personality and Social Psychology, 65,* 930–941.

Niederhoffer, K. G., & Pennebaker, J. W. (2002). Linguistic style matching in social interaction. *Journal of Language and Social Psychology, 21,* 337–360.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). Linguistic inquiry and word count (LIWC) [Computer software]. Mahwah, NJ: Erlbaum.

Pennebaker, J. W., & Graybeal, A. (2001). Patterns of natural language use: Disclosure, personality, and social integration. *Current Directions in Psychological Science, 10,* 90–93.

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology, 77,* 1296–1312.

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of language use: Our words, our selves. *Annual Review of Psychology, 54,* 547–577.

Reis, H. T., Lin, Y. -C., Bennett, M. E., & Nezlek, J. B. (1993). Change and consistency in social participation during early adulthood. *Developmental Psychology, 29,* 633–645.

Reis, H. T., & Wheeler, L. (1991). Studying social interaction with the Rochester Interaction Record. *Advances in Experimental Social Psychology, 24,* 269–318.

Robinson, J. P. (1977). *How Americans use time: A social–psychological analysis of everyday behavior.* New York: Praeger.

Robinson, J. P. (1985). The validity and reliability of diaries versus alternative time use measures. In F. T. Juster & F. P. Stafford (Eds.), *Time, goods, and well-being* (pp. 63–92). Ann Arbor, MI: Institute for Social Research.

Rozin, P. (2001). Social psychology and science: Some lessons from Solomon Asch. *Personality and Social Psychology Review, 5,* 2–14.

Sanford, F. H. (1942). Speech and personality. *Psychological Bulletin, 39,* 811–845.

Schnurr, P. P., Rosenberg, S. D., Oxman, T. E., & Tucker, G. J. (1986). A methodological note on content analysis: Estimates of reliability. *Journal of Personality Assessment, 50,* 601–609.

Schwarz, N., & Sudman, S. (Eds) (1993). *Autobiographical memory and the validity of retrospective self-reports.* New York: Springer-Verlag.

Stone, A., & Shiffman, S. (1994). Ecological momentary assessment (EMA) in behavioral medicine. *Annals of Behavioral Medicine, 16,* 199–202.

Stone, A. A., Turkkan, J. S., Bachrach, C. A., Jobe, J. B., Kurtzman, H. S., & Cain, V. S. (Eds.). (2000). *The science of self-report: Implications for research and practice.* Mahwah, NJ: Erlbaum.

Swann, W. B. (1987). Identity negotiation: Where two roads meet. *Journal of Personality and Social Psychology, 53,* 1038–1051.

Szalai, A., Converse, P. E., Feldheim, P., Scheuch, E. K., & Stone, J. P. (1972). *The use of time: Daily activities of urban and suburban population in twelve countries.* The Hague, the Netherlands: Mouton.

Tannen, D. (1990). *You just don't understand: Women and men in conversation.* New York: Morrow.

Thomson, R., Murachver, T., & Green, J. (2001). Where is the gender in gendered language? *Psychological Science, 121,* 171–175.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response.* Cambridge, England: Cambridge University Press.

Weintraub, W. (1989). *Verbal behavior in everyday life.* New York: Springer-Verlag.

Received December 21, 2001
Revision received August 9, 2002
Accepted August 22, 2002 ■