Running Head: Electronically Activated Recorder

In press, Mehl & Conner (Eds.) Handbook of Research Methods for Studying Daily Life. New York, NY: Guilford Press.

Naturalistic Observation Sampling: The Electronically Activated Recorder (EAR)

Matthias R. Mehl

Megan L. Robbins

University of Arizona

Naturalistic Observation Sampling: The Electronically Activated Recorder (EAR)

According to many laypersons, psychologists are people-watchers; what they do is observe behavior. Ironically, whereas the observation of subjects in their natural habitat, or naturalistic observation, is a fairly common method in neighboring disciplines (e.g., anthropology, sociology, primatology), it has a remarkably thin history in psychology. In this chapter, we follow Trull's recommendation of "expanding the aperture of psychological assessment" (2007; p. 1), and highlight and review one specific methodology for studying daily life that allows for the unobtrusive sampling of naturalistic observations.

Figure 1 shows a simplified method matrix. It organizes types of methods used in social science research according to whether the data collection is based on self-reports or behavioral observation and whether it takes place in the lab or in the participants' natural environments (for sake of simplicity, physiological assessments are excluded). The upper left quadrant shows the generic global/ retrospective self-report questionnaire (e.g., a standardized behavior checklist). The upper right quadrant of this matrix shows the typical self-report methods for studying daily life such as self-report based ambulatory assessment (AA; Fahrenberg, Myrtek, Pawlik, & Perrez, 2007; Wilhelm & Grossman, 2010), ecological momentary assessment (EMA; Stone & Shiffman, 1994), daily diary (Bolger, Davis, & Rafaeli, 2003) and experience sampling methods (ESM; Hektner, Schmidt, & Csikszentmihalyi, 2007). The lower left quadrant contains lab-based observational methods such as the videotaping of couple (e.g., Heyman, 2001) or family interactions (e.g., Margolin et al., 1998).

[Insert Figure 12.1 about here]

As apparent from Figure 1, the lower right quadrant, behavioral observation in the natural environment, is not well represented (for exceptions see chapters by Bussman & Ebner-Priemer,

this volume and Goodwin, this volume). In fact, in psychology, extremely few studies have employed person-centered, naturalistic observation (e.g., Barker & Wright, 1951; Craik, 2000). Funder (2007) pointed out that, among other reasons, this is because it is not straightforward how one would go about collecting truly naturalistic behavioral data. Essentially, it seems, it would require a "detective's report [that] would specify in exact detail everything the participant said and did, and with whom, in all of the contexts of the participant's life" (p. 41). Because this is ultimately impossible, self-report-based momentary assessment methods are generally considered the best available proxy to behavioral observation in the field (Conner et al., 2009).

From a multi-method perspective, however, momentary and global/ retrospective self-reports share important method variance because both derive their data from participants' reports of their introspections and perceptions–in the method matrix both are located within the same row. Therefore, some of the concerns raised for global/ retrospective self-reports potentially also apply to momentary self-reports (e.g. impression management, self-deceptive enhancement, limitations to what participants are aware of; Piasecki, Hufford, Solhan, & Trull, 2007). Thus, to complete the social science researcher's tool kit, it would be desirable to fill the lower right quadrant by complementing momentary self-report data with momentary observational data.

A Method for the Naturalistic Observation of Daily Life: The Electronically Activated Recorder (EAR)

Over the last 12 years, we have developed the *Electronically Activated Recorder or EAR* (Mehl, Pennebaker, Crow, Dabbs, & Price, 2001), a method that unobtrusively samples acoustic observations of participants' momentary environments within the natural flow of their lives.

What is the EAR?

The EAR is a portable audio recorder that is set to periodically record brief snippets of

ambient sounds. Participants wear it attached to their belts or in a purse-like bag while going about their daily lives. In tracking moment-to-moment ambient sounds around the participants, it yields acoustic logs of their days as they naturally unfold. In sampling only a fraction of the time instead of recording continuously, it makes large-scale naturalistic observation studies feasible.

Since its conception in 2001 the EAR has evolved from a modified microcassette recorder (Mehl et al., 2001; Mehl & Pennebaker, 2003b) via a micro-chip triggered digital voice recorder (e.g., Hasler, Mehl, Bootzin, & Vazire, 2008; Holtzman, Vazire, & Mehl, 2010; Mehl, 2006; Mehl & Pennebaker, 2003a; Mehl, Gosling, & Pennebaker, 2006) to today's third generation EAR system, which runs on a Personal Digital Assistant (PDA or handheld computer). The PDA-based EAR system has some critical advantages: (1) it is software-based and runs on regular, commercial devices (i.e. requires no custom-designed hardware); (2) it is available at a reasonable price (the cost of a PDA); (4) it allows for freely programmed recording schedules (e.g., 30 sec every 12.5 min, 5 min every hour) as well as blackout periods with no recordings (e.g., over night). Finally, because now the traditional, self-report-based AA methods and the EAR use the same electronic device, it is possible to merge both methodologies. Figure 2 illustrates how the PDA-based EAR system is worn by a person.

[Insert Figure 12.2 about here]

How does the EAR compare to traditional, self-report-based AA methods?

As a psychological real-time data capture method, the EAR compares most directly to self-report-based AA (or EMA) methods (Bolger, et al., 2003; Conner, et al., 2009; Stone et al., 2007). Table 1 summarizes important similarities and differences between the two methods.

[Insert Table 12.1 about here]

Both methodologies are naturalistic in their approach and based on ecological research

perspectives (Fahrenberg et al., 2007; Reis & Gosling, 2010; Wilhelm & Grossman, 2010). Whereas for traditional self-report-based AAs both paper-and-pencil and PDA-based versions are available (Hektner et al., 2007; see also Kubiak and Krog, this volume), the EAR runs only electronically. Also, using the distinction introduced by Conner and Lehman, this volume, self-report based AA data are provided actively through participants' voluntary actions (e.g., checking a box in response to an item) whereas EAR data are collected passively through automatic recordings without participants' direct involvement (other than wearing the device).

The most important difference between the two methods lies in the fact that traditional AA or EMA methods are based on momentary self-reports whereas the EAR is based on momentary behavioral observation. The two types of methods hence adopt different assessment perspectives: the self, with the corresponding subjective, experiential account, versus the bystander, or observer, with the corresponding objective (i.e. 'person as object') account.

Self-report-based AA methods by nature require participants' awareness of the assessment. In contrast, the EAR operates imperceptibly; participants never know when the recorder is on or off. Further, after an initial period of device-induced self-awareness (approximately 2 hours), participants' generally habituate to wearing the EAR and often report forgetting about it for extended periods of time (Mehl & Holleran, 2007).

The two methods further differ in the burden they place on participants. Self-report-based AA methods come with the practical burden of requiring participants to intermittently interrupt the flow of their daily lives to answer a series of questions. This practical burden creates an upper limit for the number of prompts that can be implemented per day, and the number of questions that can be asked per prompt. The practical burden of the EAR consists of wearing the device; this burden is relatively low and independent of the sampling rate or the amount of

information that is extracted (i.e. coded) from the sound files. However, the EAR places a very different burden on participants: the psychological discomfort of knowing one is intermittently recorded (sometimes referred to as evaluation apprehension). Therefore, EAR data collection is limited to sampling rates that result in privacy intrusions that are tolerable for participants.

Finally, the two methods also differ considerably in the kinds of burden they place on the researcher. With self-report-based AA methods, the researcher's challenge consists of adequately instructing and training participants to ensure high compliance and data quality. With the EAR, very little participant preparation is necessary (other than creating good rapport). Participants receive the device activated and, ideally, wear it without ever touching a button. However, with the EAR, a major challenge for the researcher lies in preparing the large amount of sound data. Somehow, the rich information contained in the sound files needs to be quantified. This usually means a sizable team of research assistants coding and transcribing sound data for hundreds of hours. Therefore, EAR data collection is often also practically limited by an investigator's lab capacity for coding the large amount of sound data.

Taken together, these practical and conceptual differences between traditional self-report-based AA methods and the EAR suggest that the two methodologies are best suited for slightly different assessments. In capturing the agent's "insider" perspective, self-report-based AA methods are optimized for the assessment of participants' subjective experiences and perceptions (e.g., thoughts, feelings, attributions). In contrast, in capturing the observer's "outsider" perspective, the EAR is optimized for assessing audible aspects of participants' objective social environments and interactions (e.g., social settings, communication behaviors, language use).

What information can be extracted from the EAR recordings?

To extract relevant information from the sampled ambient sounds, researchers can either

adopt a psychological rating or a behavior coding approach (Sillars, 1991). With the psychological rating approach, expert raters listen to the full set or selected segments of participants' sound files and judge the degree to which they indicate the presence of a construct of interest. For example, relationship experts could rate captured conversations with the participants' significant other on relationship satisfaction, social support, expressed emotions, or protective buffering (Kerig & Baucom, 2004). Or, communication experts could rate captured workplace conversations for how competent the participant appeared in them (Holleran, Whitehead, Schmader, & Mehl, in press). In these cases information is extracted at a molar, psychological level. Reliability can be determined from the consensus among the expert raters, and the construct validity of the ratings emerges from comparisons with established criterion measures (e.g., self- or spousal reports; coworker or supervisor ratings).

In our research, we have primarily worked with behavior codings. With this approach, information is extracted at the molecular level of the raw behavior. Trained coders listen to all of a participant's EAR recordings and code each sound file using a standardized coding system. Over the years, we have developed and refined the *Social Environment Coding of Sound Inventory* (SECSI; Mehl et al., 2006; Mehl & Pennebaker, 2003b) to capture acoustically-detectible aspects of participants' social environments and interactions. In its basic form, the SECSI comprises four category clusters: (1) the person's current *location* (e.g., at home, outdoors, in transit; all inferred from ambient, audible cues to their location, such as the wind blowing outside or the sound of surrounding traffic while inside a car), (2) *activity* (e.g., listening to music, watching TV, eating), (3) *interaction* (e.g., alone, talking, on the phone) and (4) *emotional expression* (e.g., laughing, crying, sighing).

[Insert Figure 12.3 about here]

Conceptually, it captures information about how individuals (1) select themselves into social environments (e.g., displaying a preference for spending time in one-on-one versus group settings) and (2) interact with their social environments (e.g., laughing or sighing a lot; see Figure 3). Adding to the basic SECSI system, we have then developed more specific coding systems that aim at capturing, for example, the topics of students' daily conversations (e.g., school, politics, entertainment, sex), coping-relevant aspects of patients' interactions with their support networks (e.g., disclosure, positive or negative support received), and behavioral residue of meditation training in daily life (e.g., gratitude, affection, empathy). Finally, in our lab, coders also routinely transcribe all of the participants' utterances captured by the EAR. We then content analyze the verbatim transcripts, usually using the Linguistic Inquiry and Word Count software (Pennebaker, Booth, & Francis, 2007), to obtain information about participants' linguistic styles.

To assess the psychometric properties of the EAR data, we obtain estimates of intercoder reliability by having all coders of a study code a standard set of training EAR recordings. Consistent with the specific, concrete, and behavioral nature of the codings (e.g., 'talking' or 'laughing'), intercoder reliabilities tend to be high. The majority of the SECSI categories have reliabilities that exceed r = .80 (Mehl et al., 2006; Mehl & Pennebaker, 2003b; Vazire & Mehl, 2008). It is an advantage of the coding approach that behavior codings at the molecular level (e.g., 'talking') are less susceptible to interpretational ambiguity than psychological ratings at the molar level (e.g., 'relationship satisfaction', 'competence'). Yet, to the extent possible (given the large amount of data), it is recommendable to have at least two independent research assistants code the data to increase reliability.

What are ethical considerations around the EAR method?

Recording ambient sounds around participants raises ethical and legal questions. EAR

studies conducted in our lab routinely implement a series of safeguards to protect participants' privacy and to ensure the confidentiality of the data (Mehl, 2007). We have found these safeguards to be highly effective at alleviating concerns that participants may have about the method. First, the EAR is programmed to record only a fraction of a person's day. Our original pattern, 30 sec every 12.5 min, recorded less than 5% of the time and left more than 95% of participants' days private in the first place, but still yielded almost an hour of audio data per day. Now we usually sample 50 sec every 9 min which still leaves 90% of the time unrecorded. Second, the recordings are kept short; 30 sec or 50 sec recordings are long enough to reliably extract basic behavioral information, yet they are short enough to capture little contextualized personal information. Finally, and most importantly, all participants can listen to their EAR recordings and delete parts they do not want on record before the investigators access the data. In one study (Mehl et al., 2006), 19 out of 96 participants (19.8%) reviewed their recordings, but only three erased sound files (10 in total). In another study with 13 arthritis patients, only 1 out of 2,948 waking sound files was erased (Robbins, Mehl, Holleran, & Kasle, in press). This suggests that participants feel generally quite comfortable with sharing the sounds of their daily lives under the safeguards that we routinely implement.

However, the more serious concerns revolve not around the participants themselves, but around bystanders who are not directly involved in the study but whose behaviors are captured by the EAR. 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 more ambiguous. In most parts of the U.S. (including Texas and Arizona where the studies from our lab have been conducted), recordings can be made legally if at least one of the interactants (e.g., the participant who is wearing the EAR) has knowledge of the recording device. A small

number of states only allow recordings if all interactants have knowledge of the recording device. Even in the most legally restrictive states, however, unauthorized recordings are only problematic if they are personally identifiable.

In EAR studies from our lab, participants are encouraged to wear the microphone visibly and to openly mention the EAR in conversations with others. Irrespective of such notification, anonymity of other people's utterances is of paramount importance, because their behavior is collected without explicit informed consent. As mentioned above, the brief recording snippets minimize the chance that personally identifying information about a third person is captured. For further protection, the sound files are coded by trained staff that is certified for research with human subjects. In the coding process, then, any personally identifying information is omitted from the transcripts. Finally, participants always have the option of erasing sound files before the researchers can access them. It is thus highly unlikely that the EAR paradigm as we have established it violates privacy rights of people who are inadvertently recorded.

How obtrusive is the method and how well do participants comply with wearing the EAR device?

The EAR method requires participants to tolerate being intermittently recorded without exactly knowing when. This can create evaluation apprehension and result in reactance (i.e. censored or artificial behavior) or non-compliance (i.e. not wearing the EAR). Thus, it is critical to estimate how obtrusive the method is in daily life and how well participants comply with it.

Mehl and Holleran (2007) addressed these questions by analyzing measures of both self-reported and behaviorally-assessed EAR obtrusiveness and compliance in two samples: a short-term (two days; Mehl et al., 2006) and a longer-term (10-11 days; Mehl & Pennebaker, 2003a) monitoring. Self-reported obtrusiveness was measured with items such as "To what degree were you generally aware of the EAR?" or "To what degree did the EAR impede on your daily

activities?" As a behavioral measure of obtrusiveness, the coders counted in how many sound files participants mentioned the EAR with others. As a self-report measure of compliance, participants reported what percentage of the day they were wearing the EAR. Finally, as a behavioral compliance measure, the coders counted the number of sound files that indicated that participants were not wearing the EAR. 'Not wearing the EAR' was coded if over a 30-sec recording period no ambient sounds at all were recorded—not even sounds of breathing or clothes rubbing against the microphone.

The analyses painted the following picture about the method's obtrusiveness: Closely after receiving the EAR, participants go through a brief period of heightened self-awareness in which conversations about the EAR are frequent. Within two hours of wearing the device, however, most participants habituate to the method and rarely mentioned it with others anymore (Panel A of Figure 4). This habituation effect was found not only for the short-term monitoring but also for the longer-term monitoring. In the longitudinal sample, some individuals initially talked about the method more than others; yet after 5-6 days of wearing the device, virtually all participants had adjusted to it and barely mentioned it anymore in their daily conversations.

[Insert Figure 4 about here]

The study further yielded the following findings about participants' compliance: In the short-term monitoring, participants' compliance was very high in the first hours after they had received the EAR. Non-compliance gradually increased over time and leveled off at about 10-12% on the second day of monitoring (Panel B of Figure 4). Compliance in the longer-term monitoring was high for at least 6 days. After that, variability in non-compliance increased suggesting that some participants' tolerance threshold may have been reached.

The compliance data reported in Mehl and Holleran (2007) are based on two studies with

undergraduate student samples. We have since run a series of EAR studies with samples of older, working adults (e.g., individuals with rheumatoid arthritis, couples in which one member was receiving treatment for breast cancer; faculty members at a Tier I research institution) and have obtained highly comparable results regarding EAR obtrusiveness and compliance.

Taken together, this suggests (1) that EAR compliance and obtrusiveness can be reliably assessed, (2) that compliance is generally high and comparable to what has been reported for self-report-based EMAs (Green, Rafaeli, Bolger, Shrout, & Reis, 2006) and (c) that after an initial habituation period of about two to three hours, the method operates fairly unobtrusively and does not impede participants much in their normal activities.

What things are captured on the sound files? To what extent does the EAR reveal real life?

As the metaphorical researcher's ear on the participant's lapel, the EAR essentially eavesdrops on people's daily lives. That naturally begs the question of what things are captured on the sound files and to what extent the EAR reveals real life. One of our first "aha!" experiences when we started doing EAR research was how ordinary and mundane real life really is. The sound files we obtained from participants first and foremost documented that for most people most of real life is not really thrilling, glittery, and extraordinary. In the end, our daily lives tend to be fairly "average". In the words of one of our participants (after listening to her own sound files), "I, as probably most people, like to think of myself as interesting and superior. Listening to myself, however, I have concluded that I am most certainly not. I am just like everybody else." Much of what the EAR captures is either silence (apart from ambient noises) or rather banal and linguistically unrefined utterances that reveal participants engaged in the pursuit of their daily activities (e.g., school, commute, watching TV). In essence, the majority of the sound files speak to the ordinary, humdrum nature of daily life (Craik, 2000).

Yet, with its fine-grained grid of observations, the EAR also regularly captures some of the less publicly presentable aspects of a person's social behavior. For example, the EAR at times catches intimate conversations as well as emotional outbursts, arguments, and profanity. In addition to documenting a person's behavior "on stage", it also reveals some of those moments where humans are caught off-guard showing their usually hidden, weak, and unpolished faces (Goffman, 1959). This potential of the EAR to capture "off-stage" behavior can, for example, be used by researchers to get a better handle on the assessment of theoretically important but methodologically notoriously difficult to measure evaluatively loaded behaviors such as small talk (Mehl et al., 2010), swearing (Robbins et al., 2010) or negative social support (Mehl, 2007).

How frequently and for how long should the EAR sample?

The first EAR studies all sampled at a rate of 30 seconds every 12.5 min. This somewhat strange sampling pattern was initially chosen to (a) obtain about 5 data points per hour, (b) avoid the oversampling of periodic behaviors (e.g., the news at the full hour), and (c)—indeed—to fit one day of monitoring onto one side of a 90 min micro-cassette! Ironically, what started largely as a product of pragmatics and good guesses turned out to be a rather effective solution. This sampling pattern has not only resulted in a number of unique findings (e.g., Mehl, Vazire, Ramirez-Esparza, Slatcher, & Pennebaker, 2007), it also turned out to be a psychometrically sound. Simulating different sampling patterns from continuous audio recordings of 11 children, Fellows, Hixon, Slatcher, and Pennebaker (2010) recently empirically confirmed that—using a given sampling rate—shorter (30 sec) recording segments are superior to longer ones (90 sec) for adequately representing the participants' full-day behavior. Further–and not surprisingly–higher sampling rates outperformed lower ones (e.g., 12.5% vs. 2.5% of the time).

In the more recent EAR studies, we have switched to sampling for 50 sec every 18 or 9

min resulting in approximately 5% or 10% of the day being recorded. We implemented this modification (a) to accommodate reviewers' concerns about more contexts seeming necessary for obtaining valid codings and (b) to capture additional context for coding constructs at a more molar, psychological level (e.g., disclosure). Because our recording segments are still relatively short and because we sample at a relatively high rate, our current pattern is well supported by Fellows et al.'s (2010) findings. Ideally, though, future research would further explore optimal sampling rates for behaviors of varying frequency and duration.

A final question concerns the duration of the EAR monitoring. In our first EAR study (Mehl & Pennebaker, 2003b), participants were wearing the EAR twice for two full days separated by approximately four weeks. The analyses revealed high stability coefficients for most behaviors suggesting that two-day monitorings are sufficient to capture habitual aspects of people's daily social environments and interactions. To be sure, longer monitoring periods are always preferable (and we have employed them in some studies). Yet, given the labor intensity of the coding, longer monitorings in reality often compete against larger samples—with, based on our experience, the sample-size advantage being stronger than the number-of-occasions one (see also Bolger and Laurenceau, this volume).

How can researchers get the EAR? What practical things are important to know to use it?

Researchers can obtain a copy of the software directly from our lab free of cost (and, so, only need to purchase a PDA device, a Secure Digital (SD) card, and a protective case to get started). All we ask for in return is for users to share their experiences and to provide feedback on how the system can be improved. Our hope is that this change will lower the psychological and economic hurdles for researchers who are interested in the EAR and foster a more widespread use of the method.

It is beyond the scope of this article to provide all the necessary practical information for running an EAR study (due to technical progress and commercial device lifecycles, this information is also subject to frequent change). However, we have maintained a researcher's guide with this purpose in mind. This guide is available from my lab and contains (1) hardware recommendations (e.g., the software runs only on Windows Mobile operated handheld computers or cell phones), (2) instructions for how to install and use the software, (3) a sample consent form, (4) a set of standard questionnaires (e.g., EAR compliance and obtrusiveness questionnaire), and (5) a script for how to administer the EAR. Apart from providing this guide, we are always happy to help "jump start" and troubleshoot.

What is the Added Value of the EAR Method for Studying Daily Life?

Conceptually, the EAR now provides a tool in the social science researcher's toolbox for person-centered behavioral observations in the natural environment. Ultimately, though, methodologies justify their existence not from filling quadrants in a method matrix, but from bringing unique potentials to psychological assessment. What, then, is the added value of the EAR method? In this section, we discuss three ways in which the EAR, as a naturalistic observation method, can uniquely inform the psychological study of daily life.

It can help calibrate psychological effects against frequencies of real-world behavior.

A persistent problem that the EAR can help resolve is the calibration and interpretation of psychological effects. In a seminal article, Sechrest, McKnight, and McKnight (1996) pointed out that "very few psychological measures of any kind are expressed in a metric that is intuitively or immediately meaningful" (p. 1065) and that the discipline would benefit from developing "a better understanding of the measures by which the phenomena with which we concern ourselves are gauged" (p. 1068). Indeed, the vast majority of standardized measures use

arbitrary metrics, such as 5- or 7-point rating scales. Such measures have no clear referents that inform about what scoring at a certain level (e.g., a "5" on optimism) means for how a person fares in important domains of life. For example, how much less time does a person with a "4" on an extraversion scale spend alone compared to a person with a "2"? And, how does a person's daily life change if an intervention reduces his depression score by 7 points? Finding answers to questions like these is crucial for understanding the implications of psychological effects. Yet, the field has struggled greatly with accomplishing this (Blanton & Jaccard, 2006; Kazdin, 2006).

One advantage of the EAR is that its sound-file based behavioral codings can be readily converted into a metric that is non-arbitrary, intuitively meaningful, and inherently real-world relevant. If the EAR captures a person talking in 40 out of 120 recordings, one can estimate that the person spent about a third of her time awake (or about 5 hours) talking. Or, if TV sounds are present in 20% of the recordings, one can estimate that the person was (actively or passively) watching TV 20% of the waking day (or about 3 ½ hours). By linking EAR-derived frequencies of daily behavior to the metrics of measures, a better understanding of the real-world implications of psychological effects can be obtained.

For example, in one EAR study (Mehl et al., 2006), Extraversion was correlated r = -.27 with time spent alone, and Conscientiousness r = .42 with time spent in class. Converted into a more meaningful metric, this suggests that participants who marked a "4" on the 5-point Extraversion scale spent almost 10% less time alone than those who marked a "2" (70.8% vs. 61.4%). And, participants who marked a "4" on the 5-point Conscientiousness scale spent about three times more time in class than those who marked a "2" (11.9% vs. 4.1%). Similarly, testing the myth that women are by a factor more verbose than men, Mehl et al. (2007) revealed, based on six EAR studies, that both men and women use about 16,000 words per day. Compared to a

range of over 46,000 words between the least and most talkative individual (695 vs. 47,016), a sex difference of 546 words rendered significance testing close to meaningless—and spoke impressively to the magnitude of individual differences. Finally, Mehl et al. (2010) recently found well-being to be related to having less small talk and more substantive conversations. The magnitudes of these effects were vividly illustrated by the fact that, compared to the unhappiest participants (-2.0 *SD*) the happiest ones (+1.5 *SD*) had roughly one third as much small talk (10.2% vs. 28.3%) and twice as many substantive conversations (45.9% vs. 21.8%).

Thus, in facilitating an absolute metric in the measurement of daily social behaviors and environments, the EAR can help "benchmark" and interpret psychological effects.

It can provide ecological, behavioral criteria that are independent of self-report.

The "criterion problem" is a vexing issue in the field (Kruglanski, 1989). How can we study how much self-insight people have into "how they really are" if the only way to assess "how they really are" is to ask them how they are? Or, more broadly, how can we study processes underlying self- and social perceptions if we do not have a way to line up these perceptions with independent measures of "actual reality"? In cases where it is necessary to measure behavioral criteria independent of self-report, the EAR can help accomplish this.

For example, Vazire and Mehl (2008) tested the accuracy of self- and other-reports by comparing the predictive validity of participants' self-ratings of how much they engage in different daily behaviors (e.g., talking on the phone, laughing, watching TV, listening to music) to similar ratings obtained from people who knew the participants well. The frequency with which the EAR captured participants actually engaging in these behaviors (e.g., actual time spent on the phone over a period of four days) served as "impartial" accuracy criterion. Self- and other-ratings showed identical validity but also uniquely predicted certain behaviors. For

example, whereas the self was better at estimating the amount of time spent arguing, friends had a more accurate picture of how sociable participants were, that is how much time they spent in the company of others. Importantly, to avoid giving one perspective an undue predictive advantage, it was critical to minimize shared method variance with both. The EAR-derived behavior counts maximally accomplished this while preserving the study's ecological focus.

Similarly, responding to Terraciano et al.'s (2005) influential finding that national stereotypes have zero validity, Heine, Buchtel, and Norenzayan (2008) argued that "comparing means on subjective Likert self-report scales is the most commonly used method for investigating cross-cultural differences, yet there are many methodological challenges associated with this approach" (p. 309). Following their advice to concentrate on behavioral trait markers, Ramirez-Esparza, Mehl, Alvarez Bermudez, and Pennebaker (2009) compared Americans' and Mexicans' sociability in a binational EAR study. They found that although American participants reported being more sociable than their Mexican counterparts, they spent less time with others and had fewer social (i.e. non-instrumental) conversations. Intriguingly, whereas Americans rated themselves significantly *higher* than Mexicans on the item "I see myself as a person who is talkative", they spent in fact almost 10% *less* time talking (34.3% vs. 43.2%). Thus, a behavior counting approach such as the one employed by the EAR can help with circumventing methodological problems around the use of self-report in cross-cultural research.

Finally, in collaboration with Dr. Raison from the Mind-Body Program at Emory University, we are currently running a randomized controlled trial using the EAR method to test how meditation training changes its practitioners' social behaviors and environments. There is consensus among meditation researchers that a clear demonstration of the real-world, prosocial effects of meditation is crucial for the field. Importantly, however, retrospective and even

momentary self-reports cannot unambiguously distinguish change at the level of the self-concept from change at the objective, behavioral level. If participants report having emerged as kinder, calmer, and more compassionate after meditation training, is it because meditation aims at and works though transforming (self-)perceptions—and therefore report—or is it because it leads practitioners to, in fact, engage in more kind, calm, and compassionate acts in their daily lives. Coding the EAR sound files for acts of gratitude, altruism, and compassion can circumvent this methodological issue and help provide the direct behavioral evidence the field has been seeking.

Taken together, in providing ecological, behavioral criteria that are independent of self-reports, the EAR can contribute in unique ways to resolving important questions in the field.

It can help with the assessment of subtle and habitual social behaviors that evade self-report

Asking participants to accurately report on subtle or habitual social behaviors is a task that often goes beyond what self-report can accomplish. For example, Schwarz (2007) demonstrated that frequent, mundane behaviors, like sighing or laughing, tend to be particularly difficult for participants to report retrospectively because occurrences become indistinguishable and irretrievable. Though self-report measures can inform about participants' self-perceptions, they often do not yield good representations of the true prevalence of subtle and habitual behavior. Multiple studies have illustrated just how precarious self-reported estimates of behavior frequencies can be (Schwarz, 2007). Momentary or end-of-day event diaries can evade memory problems associated with retrospective self-reports, but—as Piasecki et al. (2007) have pointed out—even in event diaries participants can only report what they noticed and remembered when they had to complete the diary. In the stream of our daily lives, subtle and habitual behaviors often simply don't pass the threshold of consciousness. Therefore, the study

of such subtle and habitual aspects of our daily lives necessitates a behavioral observation approach.

One study that exemplifies how the EAR can capture subtle social behaviors tested the degree to which spontaneous sighing is a behavioral indicator of depression among rheumatoid arthritis (RA) patients (Robbins, Mehl, Holleran, & Kasle, in press). Thirteen RA patients wore the EAR (recording 50 sec every 18 min) for two weekends separated by one month. Depression and physical symptoms were assessed via questionnaires. As an "an obvious exaggerated exhalation of breath" (p. 366; Keefe & Block, 1982), incidents of sighing were readily captured by the EAR and could be reliably coded from the sampled ambient sounds. Interestingly, sighing was significantly and strongly related to patients' levels of depression and non-significantly and less strongly to their reported pain and number of flare days. Because of the small sample size, the findings are preliminary in nature. Yet, they suggest that sighing can be an observable marker of depression and may be more of a depression behavior than a pain behavior among RA patients.

Other behaviors are less subtle but highly automatic and thus difficult to report. For example, swearing's habitual and non-focal nature in everyday conversations makes it difficult to self-report (Jay, 2009). In a recent study of the intra- and interpersonal consequences of swearing, we combined data from two pilot studies of 13 women with RA and 21 women with breast cancer (Robbins, Focella, Mehl, Kasle, Weihs, & Lopez, 2010). Participants wore the EAR on weekends to track their daily conversations. All sound files were transcribed and submitted to LIWC (Pennebaker et al., 2007) to determine their degree of swearing. In addition, participants completed self-reported measures of depression and emotional support at the start of the EAR weekend and several months later at the follow-up. Consistent with the idea that

swearing can repel support at the downstream expense of psychological adjustment, swearing in the presence of others, but not alone, was related to decreases in emotional support and increases in depressive symptoms over the study period. Further, decreases in emotional support mediated the effect of swearing on disease-severity-adjusted changes in depressive symptoms. Again, these effects are preliminary in nature and may well be limited to women in midlife for whom swearing violates gender and age norms. Yet, together with the sighing findings, they highlight the importance of investigating behaviors that play an important role in daily life but are often too subtle or habitual for participants to report retrospectively or in the moment.

Summary and Conclusion

The purpose of this chapter was to provide a review and discussion of a still relatively young naturalistic observational sampling method: The Electronically Activated Recorder, or EAR. As the metaphorical researcher's ear on the participant's lapel, it eavesdrops on people's daily lives and provides highly naturalistic, experientially vivid, and psychologically rich information about their moment-to-moment (acoustic) social worlds. Within the research methods for studying daily life, the EAR clearly occupies a methodological niche; it is not for everyone and everything. It is highly labor-intensive and thus requires careful deliberation as to when it should be used instead of more economic methods (e.g., experience sampling, daily diaries). However, in providing ecological behavioral measures that are independent of self-report and often beyond what self-report can capture, it can yield valuable findings that are difficult to obtain otherwise and support the field in the mission to gradually "put a bit more behavior back into the science of behavior" (Baumeister, Vohs, & Funder, 2007, p. 401).

References

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

Baumeister, R.F., Vohs, K.D., & Funder, D.C. (2007). Psychology as the science of self-reports and finger movements. *Perspectives on Psychological Science, 2*, 396–403.

Blanton, H. & Jaccard, J. (2006). Arbitrary metrics in psychology. *American Psychologist, 61,* 27–41.

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology, 54,* 579–616.

Conner, T. S., Tennen, H., Fleeson, W., & Barrett, L. F. (2009). Experience sampling methods: A modern idiographic approach to personality research. *Social and Personality Psychology Compass, 3*, 292–313.

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: Lawrence Erlbaum Associates.

Fahrenberg, J., Myrtek, M., Pawlik, K., & Perrez, M. (2007). Ambulatory assessment - monitoring behavior in daily life settings: A behavioral-scientific challenge for psychology. *European Journal of Psychological Assessment, 23,* 206–213.

Fellows, M. D., Hixon, J. G., Slatcher, R. B., & Pennebaker, J. W. (2010). Assessing optimal sampling strategies for measuring real-world behavior. *Unpublished Manuscript*.

Funder, D. C. (2007). The personality puzzle. New York, NY: W. W. Norten & Company.

Goffman, E. (1959). *The presentation of self in everyday life*. Garden City, NY: Doubleday.

Green, A.S., Rafaeli, E., Bolger, N., Shrout, P.E., & Reis, H.T. (2006). Paper or plastic? Data

equivalence in paper and electronic diaries. *Psychological Methods, 11,* 87–105.

Hasler, B., Mehl, M. R., Bootzin, R., & Vazire, S. (2008). Preliminary evidence of diurnal rhythms in everyday behaviors associated with positive affect. *Journal of Research in Personality, 42,* 1537–1546.

Heine, S. J., Buchtel, E. E., & Norenzayan, A. (2008). What do cross-national comparisons of personality traits tell us? The case of conscientiousness. *Psychological Science, 19,* 309–313.

Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (2007). *Experience sampling method: Measuring the quality of everyday life*. Thousand Oaks, CA: Sage.

Heyman, R. E. (2001). Observation of couple conflicts: Clinical assessment applications, stubborn truths, and shaky foundations. *Psychological Assessment, 13,* 5–35.

Holleran, S. E., Whitehead, J., Schmader, T., & Mehl, M. R. (in press). Talking shop and shooting the breeze: A study of workplace conversations and job disengagement among STEM faculty. *Social Psychological and Personality Science.*

Holtzman, N. S., Vazire, S., & Mehl, M. R. (2010). Sounds like a narcissist: Behavioral manifestations of narcissism in everyday life. *Journal of Research in Personality, 44,* 478-484.

Jay, T. (2009). The utility and ubiquity of taboo words. *Perspectives on Psychological Science, 4,* 153–161.

Kazdin, A. E. (2006). Arbitrary metrics: Implications for identifying evidence-based treatments. *American Psychologist, 61,* 42–49.

Keefe, F.J., & Block, A. R. (1982). Development of an observation method for assessing pain behavior in chronic low back pain patients. *Behavior Therapy, 13,* 363–375.

Kerig, P. K., & Baucom, D. H. (Eds.) (2004). *Couple observational coding systems*. Mahwah,

NJ: Erlbaum.

Kruglanski, A.W. (1989). The psychology of being "right": The problem of accuracy in social perception and cognition. *Psychological Bulletin, 106,* 395–409.

Margolin, G., Oliver, P.H., Gordis, E.B., O'Hearn, H.G., Medina, A.M., Ghosh, C.M., Morland L. (1998). The nuts and bolts of behavioral observation of marital and family interaction. *Clinical Child and Family Psychology Review, 1,* 195-–213.

Mehl, M. R. (2006). The lay assessment of subclinical depression in daily life. *Psychological Assessment, 18,* 340–345.

Mehl, M. R. (2007). Eavesdropping on health: A naturalistic observation approach for social-health research. *Social and Personality Psychology Compass, 1,* 359–380.

Mehl, M.R., Gosling, S.D., & Pennebaker, J.W. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology, 90,* 862–877.

Mehl, M. R. & Holleran, S. E. (2007). An empirical analysis of the obtrusiveness of and participants' compliance with the Electronically Activated Recorder (EAR). *European Journal of Psychological Assessment, 23,* 248–257.

Mehl, M. R. & Pennebaker, J. W., (2003a). The social dynamics of a cultural upheaval: Social interactions surrounding September 11, 2001. *Psychological Science, 14,* 579–585.

Mehl, M. R. & Pennebaker, J. W. (2003b). The sounds of social life: A psychometric analysis of students' daily social environments and natural conversations. *Journal of Personality and Social Psychology, 84,* 857–870.

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.

Mehl, M. R., Vazire, S., Holleran, S. E., & Clark, C. S. (2010). Eavesdropping on happiness: Well-being is related to having less small talk and more substantive conversations. *Psychological Science, 21,* 539–541.

Mehl, M. R., Vazire, S., Ramírez-Esparza, N., Slatcher, R. B., & Pennebaker, J. W. (2007). Are women really more talkative than men? *Science, 317,* 82.

Pennebaker, J.W., Booth, R.J., & Francis, M.E. (2007*). Linguistic Inquiry and Word Count: LIWC 2007*. Austin, TX: LIWC (www.liwc.net).

Piasecki, T. M. Hufford, M. R., Solhan, M., & Trull, T. J. (2007). Assessing clients in their natural environments with electronic diaries: Rationale, benefits, limitations, and barriers. *Psychological Assessment, 19,* 25–43.

Ramírez-Esparza, N., Mehl, M. R., Álvarez Bermúdez, J., & Pennebaker, J. W. (2009). Are Mexicans more or less sociable than Americans? Insights from a naturalistic observation study. *Journal of Research in Personality, 43,* 1–7.

Reis, H. T., & Gosling, S. D. (2010). Social psychological methods outside the laboratory. In S. T. Fiske, D. T. Gilbert, & G. Lindzey, (Eds.), *Handbook of Social Psychology* (5th ed., vol. 1, pp. 82–114). New York: Wiley.

Robbins, M. L., Mehl, M. R., Holleran, S. E., & Kasle, S. (in press). Preliminary evidence that naturalistically observed sighing predicts depression in rheumatoid arthritis patients. *Health Psychology*.

Robbins, M. L., Focella, E. S., Mehl, M. R., Kasle, S., Weihs, K. L., & Lopez, A. M. (2010). Naturalistically observed swearing predicts decreases in emotional support and increases in depressive symptoms in women coping with disease. *Manuscript under review*.

Schwarz, N. (2007). Retrospective and concurrent self-reports: The rationale for real-time data capture. In A. A. Stone, S. S. Shiffman, A. Atienza, & L. Nebeling (Eds.), *The science of real-time data capture: Self-reports in health research* (pp. 11–26). New York: Oxford University Press.

Sechrest L., McKnight, P.E., & McKnight, K.M. (1996). Calibration of measures for psychotherapy outcome studies. *American Psychologist, 51,* 1065–1071.

Sillars, A. L. (1991). *Behavioral observation*. In B. Montgomery & S. Duck (Eds.). Studying interpersonal interaction. New York: Guilford.

Stone, A. A., Shiffman, S. S., Atienza, A. & Nebeling, L. (Eds.) (2007). *The science of real-time data capture: Self-reports in health research*. New York: Oxford University Press.

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

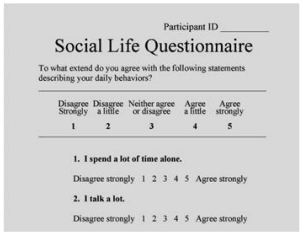
Terracciano, A., Abdel-Khalek, A. M., Ádám, N., Adamovová, L., Ahn, C.-k., Ahn, H.-n., et al., McCrae. R. R. (2005). National character does not reflect mean personality trait levels in 49 cultures. *Science, 310,* 96–100.

Trull, T. J. (2007). Expanding the aperture of psychological assessment: Introduction to the special section on innovative clinical assessment technologies and methods. *Psychological Assessment, 19,* 1–3.

Vazire, S. & Mehl, M. R. (2008). Knowing me, knowing you: The relative accuracy and unique predictive validity of self-ratings and other-ratings of daily behavior. *Journal of Personality and Social Psychology, 95,* 1202-1216.

Wilhelm, F.H., Grossman, P. (2010). Emotions beyond the laboratory: theoretical fundaments, study design, and analytic strategies for advanced ambulatory assessment. *Biological Psychology 84,* 552–569.

Figure 12.1. A simplified method matrix

Data Collection	In the Lab	In the Natural Environment
Via Self-Report	 Participant ID _____ Social Life Questionnaire To what extend do you agree with the following statements describing your daily behaviors? Disagree Strongly 1 Disagree a little 2 Neither agree or disagree 3 Agree a little 4 Agree strongly 5 1. I spend a lot of time alone. Disagree strongly 1 2 3 4 5 Agree strongly 2. I talk a lot. Disagree strongly 1 2 3 4 5 Agree strongly	
Via Behavioral Observation		?

Note: Adapted from Mehl (2007)

Figure 12.2. Picture illustrating how the PDA-based EAR-system is worn by a person.

Figure 12.3. EAR Assessment of Participants' Daily Social Environments and Interactions

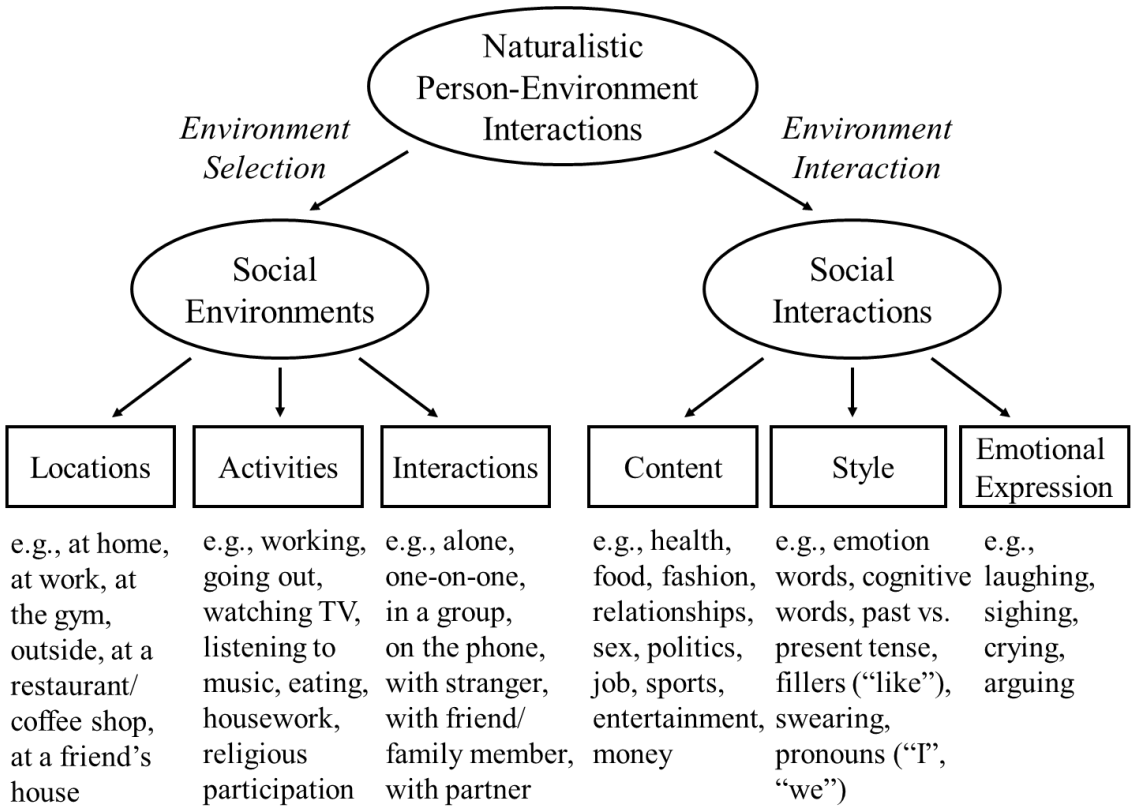
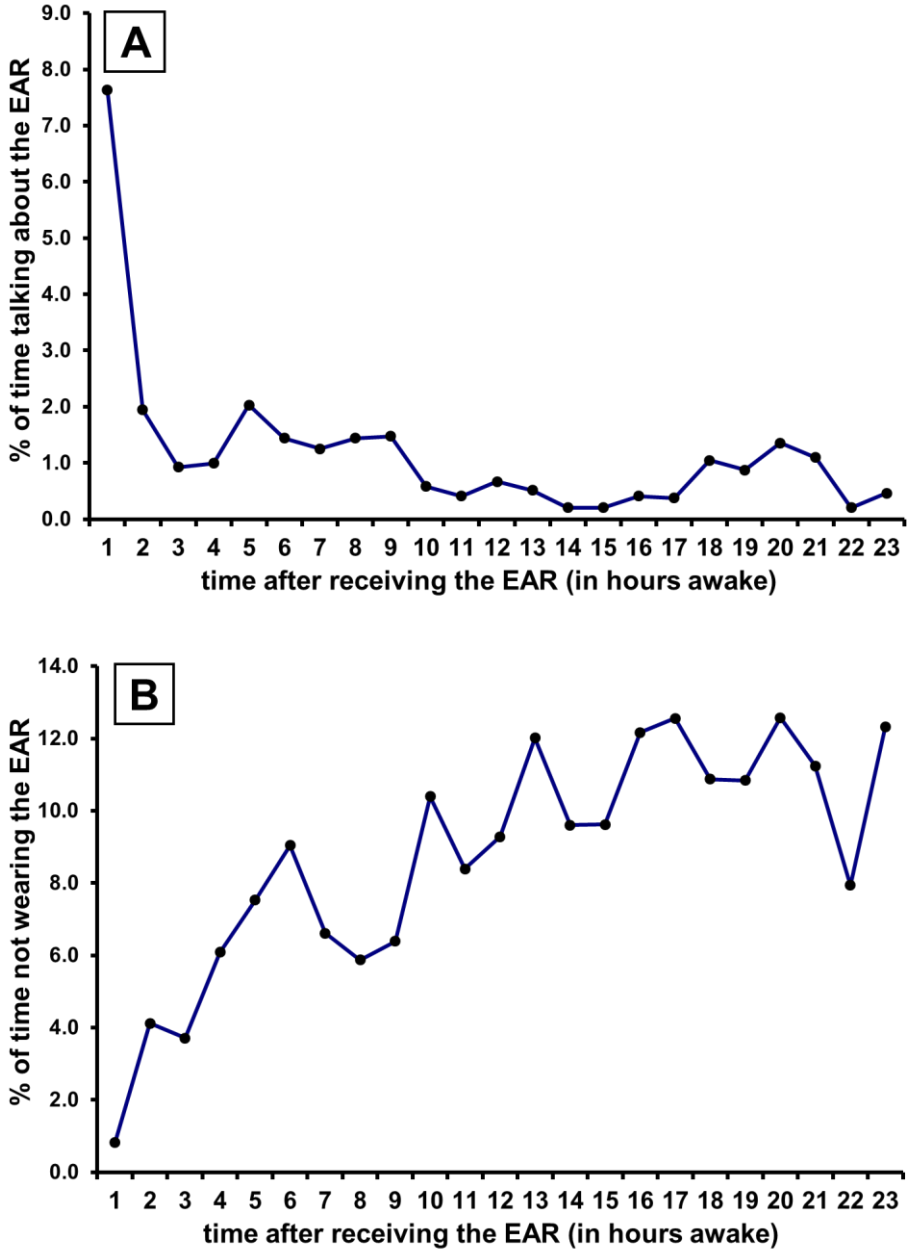


Figure 12.4. Behaviorally-Assessed Obtrusiveness of and Compliance with the EAR Method



Note: Reprinted with permission from Mehl & Holleran (2007).

Table 12.1

A Comparison between Self-report-based Ambulatory Assessment Methods and the EAR Method

	Self-report-based Ambulatory Assessment Methods	EAR-Method
Approach	Naturalistic	Naturalistic
Medium	Paper & Pencil, Electronic (PDA)	Electronic (PDA)
Mode	Active (data provided through voluntary response)	Passive (data collected through automatic recording)
Method	Self-report	Behavioral Observation
Perspective	Self (Agent)	Other (Observer)
Awareness of Assessment	High	Low after Habituation
Burden for Participant	Practical (Interruption of Daily Life)	Psychological (Intrusion of Privacy)
Burden for Researcher	Preparing Participants (Instruction and Training)	Preparing the Sound Data (Coding and Transcribing)
Data Collection Limited By	Response Burden	Privacy Considerations, Lab Capacity for Data Coding
Optimized for Assessment of	Subjective Experiences and Perceptions	Objective Social Environments and Interactions

Note: Adapted from Mehl (2007)