

# Randomly Sampling Thinking in the Natural Environment

Russell T. Hurlburt  
University of Nevada, Las Vegas

Thought sampling, the experience sampling method (ESM), and descriptive experience sampling are 3 methods of randomly sampling thinking in participants' natural environments. All use beepers that signal participants to report aspects of their experience and behavior at random intervals. Thought sampling and ESM are quantitative methods, and standard validation studies have demonstrated them to be reliable and valid. They provide insights into human experience and behavior that are difficult or impossible to attain by nonsampling methods. Descriptive experience sampling is a nonquantitative method that provides provocative descriptions of phenomena such as "unsymbolized thinking"—thoughts that occur in awareness without words, images, or any other kind of symbols—and perception that has no figure-ground phenomena.

One of the principal characteristics of the psychology of the last quarter of the 20th century has been the emphasis on thoughts as explanatory concepts, and yet psychologists remain quite ignorant about the characteristics of thoughts themselves. Although a complete exploration of why this ignorance has been tolerated is outside the scope of this article, two justifications are heard frequently: (a) Reports of subjective experience are unreliable and therefore invalid, and (b) the characteristics of thinking are already well known because people engage in thinking almost constantly. Justification (b) is accompanied by two (usually unstated) corollaries: (b1) The characteristics of the thoughts of all people are similar, and (b2) thoughts are in words. However, on the basis of the investigations described in this article (among others), I can now say with great confidence that none of these justifications are true.

Unfortunately, the widespread acceptance of these justifications has contributed to a general reluctance to conduct research that directly examines the nature and content of thinking, despite the prominence of thinking in psychological theorizing. However, a small number of researchers have attempted this direct examination of thinking by randomly sampling thoughts in participants' natural environments. I first describe two such methods that seek to quantify thinking as it naturally exists "in flight": thought sampling (or thought and mood sampling) and the experience sampling method (or ESM). Then I show that these sampling methods are reliable and valid, that they characterize experience in unique ways, and that they produce results that are difficult or impossible to obtain by other methods. Last, I describe a nonquantitative sampling method called descriptive experience sampling.

## Two Quantitative Sampling Methods

There are two loose categories of quantitative methods of sampling thinking, generally referred to as thought sampling

and the ESM. In both methods, participants are generally asked to carry beepers into their naturally occurring environments, where beeps will then occur at random or quasirandom intervals. On hearing the beep, the participant focuses on the experience (the momentary thought as well as other aspects of inner and outer experience and behavior) that was occurring at the moment of the beep and responds by filling out a quantitative questionnaire, writing a short narrative description of the experience, or some combination thereof.

The core of both methods is the attempt to provide "ecologically valid" characterizations of thinking and behavior—characterizations that mirror the naturally occurring features of people's experience as they move through their naturally occurring environments. To be ecologically valid, the "situational circumstances [of an investigation] should be made to represent, by sampling or related devices, the general or specific conditions under which the organism studied has to function" (Brunswick, 1952, p. 30).

Ecologically valid studies are rare. For example, the ideal experiment is ecologically invalid because it attempts to provide one condition that is as identical as possible for all participants. Even if that one condition is quite similar to one "under which the organism studied has to function" (which, unfortunately, is often not the case), it is still only one such condition and thus cannot possibly reflect the range of the participant's naturally occurring situations. Perhaps because psychological science is so dominated by experimentation, it has numbed itself to the risks associated with ecologically invalid investigations.

Both thought-sampling and ESM studies seek to attain ecological validity by randomly sampling participants as they move through their natural environments (Hormuth, 1986, pp. 264–265). This, of course, sacrifices the control over participants' activities that characterizes experiments, but the lack of control may be outweighed by the advantages of obtaining uncontrived samples of behavior from "real life."

## Thought Sampling

Thought-sampling (or thought-and-mood-sampling) investigations seek primarily to quantify characteristics or aspects of thinking (or of thinking and mood). The first systematic random

---

I thank Chris Heavey, Eric Klinger, and Philippe Delespaul for helpful comments on earlier drafts of this article.

Correspondence concerning this article should be addressed to Russell T. Hurlburt, Department of Psychology, University of Nevada, Las Vegas, 4505 Maryland Parkway, Las Vegas, Nevada 89154-5030. Electronic mail may be sent via Internet to russ@nevada.edu.

sampling of inner experiences in natural environments was conducted by Hurlburt (1976) and Klinger (1978–1979), both studies beginning (independently of each other) in 1974. At each randomly occurring beep, participants either wrote down a narrative description of the thought that was occurring to them at the moment (in which case, external raters subsequently rated such thoughts on a number of dimensions) or immediately rated their own thoughts on rating scales supplied by the researcher, or both.

A typical thought-sampling study is by Klinger and Cox (1987–1988). They asked participants “to use [a random] beeper for as much of each reporting day as they comfortably could” (p. 110) and to respond to each beep by filling out four short questionnaires: a Thought-Sampling Questionnaire, an Activity Report Questionnaire, a state measure of the State–Trait Anxiety Inventory, and a state measure of the Depression Adjective Checklist Form. The Thought-Sampling Questionnaire asked for a brief written report of the participant’s most recent mental content and then asked the participant to rate that thought content on 24 scales. Ratings were on 4-point scales, such as specificity (i.e., not vague), directed (i.e., requiring effort), detail, visualness, color, movement, auditoriness, interior monologue, and so on.

The 29 participants in Klinger and Cox’s (1987–1988) study provided 1,425 samples, an average of 49.1 samples per participant. These 1,425 rating vectors were factor analyzed, producing 8 factors that subsequent research has shown to be stable and reproducible. Klinger and Cox reported that most samples contained some interior monologue and that visual imagery was a more prominent characteristic than was auditory imagery for nearly all participants.

### *Experience Sampling Method*

ESM, a procedure developed in 1975 (independently of Hurlburt and Klinger) by Csikszentmihalyi and his colleagues, was designed “to study the subjective experience of persons interacting in natural environments” (Csikszentmihalyi & Larson, 1987, p. 526; see also Prescott, Csikszentmihalyi, & Graef, 1981). ESM investigations, like thought-sampling studies, beep participants at random or quasirandom intervals and ask them to report a variety of features of inner experiences. Unlike thought-sampling studies, ESM investigations also systematically ask participants to report the contexts in which that experience occurred. Such contexts include social aspects (e.g., whether the participant was alone or with how many other individuals) and environmental aspects (where the participant was at the moment of the beep, in what kind of activity was the participant engaged, etc.).

ESM differs from thought sampling in two important ways. First, ESM is a general method of investigation that may or may not examine thinking; in fact, the majority of ESM studies do not measure thinking variables; instead they are concerned with moods, quality of life variables, and so on. Second, ESM typically focuses on the context of experience (asking the participant to report where he or she was, with whom, engaged in what activity, etc.), whereas thought sampling typically focuses directly on thoughts (or thoughts and moods) themselves. Thus, ESM studies are likely to answer questions about the context of thinking (e.g., What social situations impact thinking vari-

ables?), whereas thought-sampling studies typically answer questions about the structure of thinking (e.g., What are the intercorrelations between thinking variables?). As a result, ESM studies typically collect data on only a few (if any) cognitive variables but many context variables, whereas thought-sampling studies typically collect data on many cognitive variables but only a few (if any) context variables.

A typical ESM study is by de Vries, Dijkman-Caes, and Delespaul (1990), who sampled highly anxious individuals. They divided their anxious participants into two groups: those who were also highly depressed and those who were also moderately depressed. Participants carried a programmable wrist-watch that signaled them on a preprogrammed, quasirandom schedule with 10 beeps per day between 7:30 a.m. and 11:00 p.m. At each beep, participants were to fill out an Experience Sampling Form that asked them to answer “What were you thinking at the moment of the beep?” and then to rate this thought on four 7-point Likert-type scales (pleasant, clear, normal, and distorted). Participants also rated their thoughts on three additional categorical variables: congruence of thought and activity (either yes or no), thought pathology (selecting one from a list of options, e.g., focused thinking, daydreaming, worrying, or preoccupation), and thought content (selecting one from a list of options, e.g., leisure, travel, self care, friends, sex, aggression, or nothing).

This Experience Sampling Form also asked for 7-point Likert ratings on six mood scales (cheerful, secure, social, relaxed, calm, and friendly), four activity motivation scales (like to do this, active, in control, and can concentrate on this), and three physical concern variables (hungry, tired, and not feeling well). Furthermore, the Experience Sampling Form asked for ratings of the context of the moment on three categorical variables: “Where are you?” (selecting one of the following options: home, family, network, health care, and public spaces), “What are you doing?” (selecting one from the following: nothing, self care, household, work, leisure, and travel), and “Who are you with?” (selecting one from the following: alone, family, friends, colleagues, and strangers). These 23 variables are used in most ESM studies.

For this particular study, which focused on anxiety, the Experience Sampling Form also had five additional Likert-type variables that described anxiety: “Did you feel short of breath, choking?” “Did you have palpitations, pain on the chest?” “Did you feel weak, dizzy, unsteady?” “Did you feel unreal?” “Were you afraid to die, to go crazy, to lose control?”

de Vries et al. (1990) found that there was less congruence between thought and activity when the anxious individuals were away from home than at home and that their thoughts were more disorganized when they were with other people (both these results are in marked contrast to ESM results for patients with schizophrenia who are more disorganized when they are alone; de Vries & Delespaul, 1992). Furthermore, they found that the anxious and highly depressed group “had more idle thoughts, ruminated more, experienced less focused thoughts in public, and registered most self-care and less care and involvement with others” (p. 721) than did the anxious and moderately depressed group.

### *Reliability and Validity of Sampling Results*

There is a prejudice, explainable in primarily historical terms, that maintains that it is impossible to validate reports of inner

experience because such events are fundamentally viewable by only one person. However, psychologists routinely validate constructs that are viewable by no one at all (most personality constructs, e.g.). In fact, the reliability and validity of sampling methods have been studied in ways very similar to the reliability and validity of other scientific constructs (Klinger, 1978–1979).

### *Reliability*

The reliability of psychological methods in general is explored in a variety of ways: compliance, interrater reliability, test–retest reliability, and so on. Sampling methods have been subjected to the same kinds of reliability evaluations; a few such results follow.

The basic compliance of participants with sampling methods has been investigated, for example, by Hormuth (1986), who used a quasirandom schedule of beeps. By comparing the known times of the 5,145 total beeps with the times recorded by the participants on their questionnaires, he found that participants responded to 50% within 1 min, 75% within 4 min; and 90% within 18 min, indicating that participants for the most part participated adequately. Csikszentmihalyi and Larson (1987) reviewed several studies and reached similar conclusions. However, a few sampling participants found the procedure intrusive; for example, Hnatiuk (1991) reported that many of her older participants declined to carry the beeper with them when away from home. Most of the studies of compliance compare actual beep times (as programmed directly by the experimenter) with self-reported response times. One way to eliminate the possible bias of this self-report is to use palm-top computers, with internal clocks, to record the times of the responses automatically (e.g., Shiffman et al., 1994).

Interrater reliability was examined by Hurlburt (1979). Participants wrote narrative descriptions of the thought that was occurring at the moment of the beep; these narratives were subsequently rated by two independent raters on six dimensions. The interrater reliabilities for these scales ranged from .66 to .89. Kraan et al. (1992) and Csikszentmihalyi and Larson (1987) found similar results with different scales.

The sampling analog of test–retest reliability is to compare samples from two sampling periods. Csikszentmihalyi and Larson's (1987) review demonstrated that the means of individuals' ratings from the beginning of a week were in general strongly correlated with the means of ratings from the end of the week. Such results have been found with a variety of populations, for example, older people (Hnatiuk, 1991), adolescents (Freeman, Csikszentmihalyi, & Larson, 1986), and people with schizophrenia (Delespaul, 1995).

In summary, the reliability of the sampling results has been repeatedly established by a wide variety of the same kinds of procedures used with traditional methods.

### *Validity*

Researchers have subjected the sampling strategies to many validity-evaluation procedures. Csikszentmihalyi and Larson (1987) reviewed criterion-validity studies, concluding that ESM reports of thinking do in fact relate to external criteria in expected directions. For example, thoughts by people with schizophrenia are rated as being disordered more frequently than are

thoughts by other patients (de Vries, Delespaul, Dijkman, & Theunissen, 1986), and women with bulimia engage in food-related behavior and thinking much more frequently than do women without bulimia (Johnson & Larson, 1982). Delespaul (1995, ch. 6) provided a more recent discussion of ESM criterion validity with similar conclusions.

Concurrent validity has also been demonstrated. Dijkman-Caes and Delespaul (1995) compared ESM results with structured diary reports of people with schizophrenia, finding in general high agreement. Delespaul, Reis, and de Vries (1995) compared ESM results with event-triggered recording and found significant agreement.

### *Limitations on Reliability and Validity*

In summary, the reliability and validity of the sampling results has been repeatedly demonstrated by a wide variety of the same kinds of procedures used with traditional methods. However, assessing the reliability and validity of sampling methods presents challenges not always found in traditional methods. It must be acknowledged that most of the reliability studies focus on the judgments made by the (perhaps external) raters, not on the accuracy of the recording of the thoughts themselves. This is analogous to the fact that the measurement of the reliability of, say, a depression inventory is actually a study of the measurement scale, not of the accuracy of the rating of the depression itself. Furthermore, the test–retest reliability found in studies such as those reviewed by Csikszentmihalyi and Larson (1987), which demonstrate that the means of individuals' ratings from the beginning of a week were correlated with the means from the end of the week, inherently confound reliability with trans-situational and transtemporal stability. Conventional studies seek to reduce the effect of such confounds by making measurements in situations as identical as possible (thus reducing trans-situational variability) and by measuring traits that are relatively stable over time (thus reducing transtemporal variability), but sampling studies that seek ecological validity by measuring characteristics of thoughts in natural environments cannot exercise such control. Therefore, the trans-situational and transtemporal variabilities in sampling studies should be expected to be larger than those in conventional studies, and the attainable test–retest reliability coefficients should be expected to be correspondingly smaller.

Most of the criterion validity studies involving sampling have been attempts at validating the sampling methods using conventional methods as the criterion, clearly the appropriate strategy for the newer and rarer sampling methods. However, Delespaul (1995) concluded that an adequate number of sampling validity studies had been performed so that particularly in those situations that evaluate experience in context, one should “switch from a position where the external criterion assesses ESM validity to a position where ESM is the criterion to determine the validity of the other assessment strategies” (p. 138).

### *Sampling Thoughts Provides Unique Insights*

There is thus a substantial literature demonstrating adequate reliability and validity of sampling results. I turn now to show that sampling studies are useful because they provide insights that are difficult or impossible to obtain by other methods.

An early sampling case example provides two examples of unique sampling-based observations. A 48-year-old man who I call here "Donald" (Hurlburt, 1976; Hurlburt & Sippelle, 1978) had been married for 6 years and had three young children. He described his current life situation as "better than ever": He loved his wife and adored his children, and his work circumstances were secure and rewarding. Unfortunately, he also had "free-floating" anxiety attacks so severe that he had discontinued driving his car because he was fearful that the anxiety would overtake him and cause him to black out. Clinical interviews revealed no cause for this anxiety.

Donald was asked to sample his thoughts randomly in his natural environment as he performed his normal, everyday activities, revealing that about one third of his thoughts reflected annoyance with his children (e.g., "I wish my son would stop crying" and "He left the record player on again"). Such a large percentage is quite unusual for any content category and was particularly noteworthy because Donald had no conscious notion that he was frequently annoyed with his children. Furthermore, he did not realize that this very thought sample had included such a large number of annoyance thoughts, even though he had himself just typed out a transcript of his entire series of thoughts.

Once the high frequency of annoyance thoughts was pointed out to him, he easily (although with considerable surprise) accepted that he was in fact often annoyed with his children. However, he believed that anger at his children was sinful and felt unfit as a father for having such thoughts and feelings. Donald entered into brief therapy that focused on the normality of being annoyed by one's children and on the important distinction between being annoyed and acting out aggressively. Almost immediately, his anxiety attacks disappeared.

Donald's case illustrates two of the ways in which randomly sampling thoughts can provide unique insights. First, it raises fundamental questions about what it means to say that something is "unconscious" or "outside of awareness." Suppose that Donald had been treated in a conventional (nonsampling) manner and during the course of therapy came to realize that he was frequently annoyed with his children. Both Donald and the therapist likely would have concluded that his annoyance with his children had previously been unconscious or outside of awareness. However, the sampling procedure showed definitively that this was not the case. Donald was clearly not actually unconscious of his annoyance at the moment it was occurring—for one third of his beeps, he freely identified himself as being annoyed. Instead, he characteristically neglected his annoyance when asked to retrospect on his feelings about his children. This important distinction between actually being unconscious and being characteristically neglectful during retrospection is made particularly salient by the sampling procedure.

Second, Donald's experience also suggests how random sampling of thoughts may test important hypotheses about psychopathology. A cognitive theorist might hold that the aim of therapy is to change the frequency of negative thoughts: When one reduces that frequency, the client improves. That theory might lead one to expect that Donald's decrease in anxiety would have been accompanied by a decrease in the frequency of his annoyance thoughts. The sampling method allowed a direct test of such a hypothesis by asking Donald to sample again after his anxiety had lessened. Contrary to the cognitive hypothesis,

this second sample produced approximately the same frequency of annoyance thoughts as did the first sample. For my purposes, the important point is that in some cases, sampling provides the most direct test available of a cognitive hypothesis.

Now consider other examples where sampling research provides insights that would be very difficult, if not impossible, to obtain by other methods.

Csikszentmihalyi and Figurski (1982) examined the relationship between thoughts about the self and those about mood in an ESM study with over 100 adults. They found that

thinking about [the] self is a particularly unpleasant experience. Self-thoughts are ranked lowest in involvement . . . and tied with work-thoughts for being lowest in affect. Thinking about [the] self was also associated with lower levels of activation than almost all other thought categories. (p. 20)

Contrast that finding with those of conventional studies reviewed by Ingram (1990), who concluded that individuals with psychopathological problems are on average more self-absorbed than individuals without them. The conventional studies reviewed by Ingram showed that across participants, unpleasant affect (i.e., psychopathology) is correlated with self-absorption, but those studies cannot show (as Csikszentmihalyi and Figurski's sampling study did) that within individuals, momentarily occurring self-absorbed thoughts are on average accompanied by particularly unpleasant affect at that same moment. The sampling result that self-absorbed thoughts and unpleasant affect fluctuate together over the course of a day is vitally important to understanding the relationship between self-absorption and psychopathology.

Csikszentmihalyi and Figurski (1982) also found that

emotional experience is related to self-awareness conditionally, depending upon the perceived voluntariness. During activity felt to be voluntary, people experience less affect, activation, and involvement when they are thinking about themselves than when they are thinking about something else. No such relation between emotional experience and self-awareness exists for obligatory activity. (p. 24)

This is another example of the kind of result that can be teased out of ESM data but that would be difficult, if not impossible, to obtain using retrospective measures.

Delespaul (1995) extensively studied inpatients with schizophrenia, and his sampling results provided a finer grained analysis than could conventional studies. He found, for example, that the presence of hallucinations was strongly related to the simultaneous occurrence of anxiety. Furthermore, the sampling methodology was able to show that anxiety typically rose just prior to auditory hallucinations but not prior to visual hallucinations and that anxiety lessened somewhat at the height of visual hallucinations but not at the height of auditory hallucinations.

Delespaul (1995) also found by sampling his group of patients with schizophrenia that approximately the same number of auditory hallucinations occurred when patients were alone as occurred when they were not. This might be interpreted to mean that social context (alone vs. not alone) does not influence the occurrence of hallucinations. However, when he analyzed the same data within each patient individually, he found that half his patients with schizophrenia had significantly more intense hallucinations while alone whereas one third had significantly more intense hallucinations in social situations. Only one sixth

of the patients had essentially no difference in frequency of hallucinations when alone versus when not alone. Thus, the sampling method, with its repeated-sample richness of individual data, was able to show that only one sixth of the patients with schizophrenia showed the same pattern as did the group with schizophrenia as a whole.

This ability to obtain repeated samples allowed van der Poel and Delespaul (1992) to contrast the results of repeated conventional clinical interviews with the results of repeated ESM sampling for a patient with chronic schizophrenia. The conventional clinical interviews indicated that the patient was decompensating rapidly whereas the ESM sampling results indicated an improvement in the patient's cognitive state. The ESM results proved to be the more accurate representation of the patient's actual condition and were thus useful in avoiding an unnecessary hospitalization.

Kroll-Mensing (1992) sampled both undergraduates who were anxious and depressed, finding in a between-subjects analysis that level of anxiety as measured by the Beck Anxiety Inventory (Beck & Steer, 1990) was significantly (but modestly) positively correlated with the sampled frequency of thoughts about the future, as predicted by theory. The repeated-observation nature of the sampling study allowed a within-subject analysis that found a similar relationship: Anxiety and future-oriented thoughts were often experienced at the same moment. However, the situation regarding depression was not so clear. A between-subjects analysis found that depression as measured by the Beck Depression Inventory was significantly positively correlated with frequency of thoughts about the past, as theory would predict. However, the sampling-enabled within-subjects analysis did not find that depression and past-oriented thoughts were generally experienced at the same moment.

Hunt and Rosen (1981) used a thought-sampling procedure to investigate the common conception that many chronic dieters are obsessed with thoughts of food. They sampled women who were obese and normal weight every hour for 3 days, asking the women to write down the thoughts that were occurring when they were beeped. These thoughts were subsequently blindly categorized as being related to food or nonfood. Contrary to common expectation, there was no significant difference in frequency of food-related thoughts for women who were obese and for those who were normal weight (8% and 7%, respectively).

Klinger (1984) used the thought-sampling technique to explore the cognitive-interference theory of test anxiety, which holds that test anxiety causes task-irrelevant thoughts to occur in the testing situation and that these task-irrelevant thoughts then interfere with test performance. Klinger sampled students' thinking during essay examinations but, contrary to cognitive-interference theory, did not find more task-irrelevant thoughts in test-anxious students' samples than in nonanxious students' samples. Thus, the sampling results led him to reject the cognitive-interference notion that test anxiety causes poor performance and to conclude instead that poor performance causes test anxiety.

Sampling results and conventional results should be expected to differ for one main reason beyond the issue of ecological validity discussed above: Sampling studies ask participants to report about events that are happening at the moment (or within the last few moments) and are thus not retrospective (or at least not nearly as retrospective) as are conventional techniques, such

as interviews and questionnaires that ask participants retrospectively to describe the characteristics of their thinking. There is now considerable evidence to show that human beings are not skilled at retrospectively characterizing their thinking; a few examples follow.

One such example is illustrated by the case of Donald described above (Hurlburt & Sippelle, 1978). Donald retrospectively characterized himself as having few annoyance thoughts; when in fact such thoughts were frequent. Another example is Hurlburt (1979), who beeped undergraduate students in their natural environments with the instruction to write down in a small notebook the thought and behavior that was occurring at the instant the beep began. Debriefing revealed examples of the limitations of retrospection:

Two subjects during their individual debriefings naively told the experimenter that the sample was not representative of their usual thought patterns "because I looked back over my notebook and there are no sex thoughts here; I know I spend 30 to 40% of my time thinking about sex, but the box never caught me!" (p. 108)

These students retrospectively greatly exaggerated the frequency of their sexual thoughts.

There is only one study that contrasts sampling and retrospective reports of thinking directly. Hurlburt and Melancon (1987) created questionnaires that could be used in both sampling and retrospective environments and concluded that the retrospective nature of questionnaires is likely to distort the frequency of recollections in thinking and that the indeterminate nature of retrospective questionnaires might lead to the erroneous conclusion that clarity in thinking is associated with pleasant thoughts. There are no studies that contrast long-term retrospection and sampling of thoughts directly, but contrasts between the long-term retrospection and sampling of mood do exist. For example, Freeman et al. (1986) found that adolescents' emotional states were quite stable over sampling periods spaced 2 years apart. However, these same adolescents were asked, after the second sampling period, to fill out a questionnaire that asked them to assess (retrospectively) changes in their experience over the 2-year interval. "While there was virtually no difference in the [sampled] quality of immediate experience over the 2-year span, respondents believed a strongly positive change to have taken place" (p. 175). Retrospected mood and sampled mood are distinctly different.

These examples illustrate some limitations of retrospective reporting: People often make substantial errors when they are asked to look back over some (often indeterminately specified) period of time and retrospectively characterize their experience. There are also, of course, limitations inherent in sampling studies: They do not generally control situations and conditions, do not generally manipulate conditions, are unlikely to discover relationships between rare but theoretically important events; and so on.

#### Descriptive Experience Sampling: A Provocative Variant

As seen, thought sampling and the ESM both provide quantitative measures of inner experience. Hurlburt and his colleagues have provided a nonquantitative sampling variant that they call *descriptive experience sampling*, which seeks simply to describe

inner experience rather than to quantify it based on the view that careful descriptions should be the foundation on which subsequent quantification should be built. The first application of the method was a master's thesis by Saltman (1983) under Hurlburt's supervision; a complete description of the method was provided and summarized by Hurlburt (1990, 1993, respectively):

The subject's task is to "freeze" her ongoing experience and to write a brief description of it in a notebook. . . . We . . . are not particularly interested in explanations of why they are thinking or doing what they are doing. . . . We simply wish them to describe that single experience as it naturally occurred. . . . After subjects have collected six or eight samples, they meet with us for an extended conversation about those samples. . . . This sample and discussion process is then repeated the next day . . . and is repeated again until we think we have obtained an adequate number of samples. . . . At the conclusion of the sampling period, we identify the salient characteristics of the complete set of samples. (1993, pp. 10–13)

The descriptive experience sampling method differs from both thought sampling and ESM in that descriptive experience sampling provides qualitative descriptions, not quantitative analysis. It encourages participants to develop their own descriptive language and to report aspects of their inner experiences as they themselves experience them and does not ask them to rate or categorize experiences according to predefined dimensions or categories. It is the descriptive-experience-sampling researcher's task to attempt to identify salient regularities that may occur within one participant's own idiographic series of reports and then, later, to identify nomothetic regularities that may occur between those who share demographic similarities (e.g., similar diagnostic labels).

Descriptive experience sampling must be regarded as less well established than thought sampling or ESM. Reliability and validity are more difficult (although not impossible) to establish because of the descriptive experience sampling's nonquantitative nature, and to date the only descriptive sampling studies have been performed by one group of researchers (Hurlburt and his colleagues). That being said, the descriptive results are sufficiently provocative to bear notice.

As an illustration of descriptive experience sampling results, consider the case of "Fran," a woman diagnosed as having a borderline personality (Hurlburt, 1993). Hurlburt described many salient characteristics of Fran's inner experiences, of which I discuss three. First, Hurlburt reported that Fran's inner experience was frequently populated by multiple (as many as 5 or 10) visual images, all occurring simultaneously and in the same "visual space" (i.e., these images were not a side-by-side collage but were instead all viewed straight ahead in a physically impossible overlaying that somehow did not provide any confusion for Fran herself). Fran's case is thus an example of the extreme complexity that inner experience can attain as reported by the descriptive experience sampling method. Such complexity cannot possibly be reported by any method other than sampling. For example, had Fran used a think-aloud technique—the most detailed nonsampling method—she simply could not have had time to report adequately one image, to say nothing of 5 or 10 simultaneous images.

Second, Hurlburt (1993) reported that some of Fran's visual

images (usually those with extremely negative content) often lasted for hours or days, nonstop, uninterrupted. (By contrast, the descriptive experience sampling method finds that images in healthy participants last for only a moment.) For example, Fran reported a visual and auditory image of her father "telling her off." In this image, Fran was seated at the dining room table. Her father was standing over her, pointing his finger at her, telling her she was "no good—a failure." Her mother was seen at the kitchen sink in the background looking over her shoulder at Fran. This image appeared in several successive samples, with the description being the same at each sample and apparently continued uninterrupted during the time in between for a total of at least several hours (pp. 202–205). This long-duration-image phenomenon might be considered impossible without sampling evidence.

Third, Hurlburt (1993) reported that Fran had no *figure-ground phenomenon* in either her inner image perception or her external perception—she took in an entire visual scene without focusing on any of its aspects. This conclusion was based on the fact that in repeated descriptive experience sampling interviews, Fran consistently denied the occurrence of phenomena associated with figure and ground: No part of an image appeared to be "closer" or "in better focus," and when she shifted her gaze from one image (or external object) to another, she had no experience of "zeroing in" or of the previous center of attention, "losing focus."

A major issue is of course whether Hurlburt's (1993) descriptive-experience-sampling reports about Fran accurately reflect Fran's inner experience: Fran was clearly the only person in a position to know that experience. Direct reliability studies are therefore impossible, so reliability must be indirectly inferred from validity considerations. Furthermore, one cannot apply standard validity-checking procedures (which intrinsically use across-group measures) to the idiographic observations of a single person; instead, one must infer validity idiographically, considering the unique characteristics of the particular description. I can identify five such idiographic validity considerations regarding the case of Fran.

First, the question of idiographic validity applies not to the descriptive sampling method per se but to the particular individuals who apply the method. In Fran's case, I was the researcher (Hurlburt, 1993). I might be expected to be a valid applier of the method because my previous descriptions of different people differ dramatically from each other, are sometimes surprising even to me myself, and are in agreement with other observers in those cases where more than one observer have sampled jointly (Hurlburt, 1993).

Second, the lack of figure-ground phenomenon in inner experience leads to an obvious but risky prediction that if Fran viewed the classical ambiguous figures such as the faces-vase or Jastrow's duck-rabbit, they would not "alternate" in her experience. I (Hurlburt, 1993) performed this informal validity experiment and found that Fran did in fact see both aspects of each drawing simultaneously with no alternation. A correct risky prediction can be taken as support for an underlying proposition (Popper, 1963) and therefore here as evidence of validity.

Third, I (Hurlburt, 1993) ruled out miscommunication, misunderstanding, or language deficit as alternative explanations of her failure to report figure-ground experience as follows. Fran asked to borrow the ambiguous figures to show to her coworkers,

believing that I was mistaken about the existence of the alternation phenomenon. She telephoned me a few hours later to report that to her surprise, her coworkers did in fact report the experience of alternation. In this conversation, she gave an accurate description of her coworkers' alternating experiences but still denied that such alternation occurred for her. Thus, it seemed clear that Fran understood what figure-ground phenomena are and was capable of describing them if they had existed for her.

Fourth, the descriptive-experience-sampling descriptions of Fran's inner experiences provided plausible explanations of two characteristics of her external behavior. First, during Fran's discovery of her coworkers' figure-ground phenomenon, the coworkers came to realize, much to their surprise, that Fran could pay attention to many aspects of one thing or many different things simultaneously (e.g., her frequent multiple images), as had been discovered by the descriptive experience sampling. The coworkers observed that this multiple-attention ability explained a trait that angered them all: They worked in a bank, and a frequent task was counting money. Each person would stand at a counter and count his or her own individual stacks of bills. Fran irritated her coworkers by repeatedly initiating conversations while counting, causing them to lose count. The simultaneous tasks of counting and conversing were impossible for her coworkers but simple for Fran. Thus, it seemed clear to me that the multiple-experience characteristic of Fran's inner world had real ramifications in Fran's exterior, everyday world.

The second sampling-based plausible explanation of external behavior came from Fran's psychotherapist. Before Fran had become involved in the sampling study, her psychotherapist had responded to her complaints of being preoccupied with negative thoughts by training her in *thought substitution*—a cognitive-therapeutic technique aimed at teaching her to think about something positive, based on the rationale that increasing her frequency of positive thoughts would lower the frequency of negative thoughts. However, that therapeutic intervention proved unsuccessful; sampling provided the plausible explanation that Fran was quite capable of thinking about something positive without ceasing to think about something negative.

Fifth, changes in external behavior were reflected in changes in inner experience. Near the end of the sampling, Fran experienced a remarkable improvement in her borderline symptoms: Her exterior disorganization and chaotic psychological fragility vanished. Samples obtained after this improvement were now much less complex and included the experience of figure-ground phenomena.

Taken together, these observations led me (Hurlburt, 1993) to conclude that the idiographic descriptions of Fran were indeed valid. If their validity is at least tentatively accepted, they are extremely provocative; for example, to my knowledge, no reports of visual perception without figure-ground phenomenon appear in the perception literature, and no mention is made of the possible connection of the lack of figure-ground to psychopathology.

Certainly one must be skeptical of the descriptive sampling method results, at least until they are corroborated by independent laboratories. But to dismiss the results merely as being "introspective" may be a mistake because I (Hurlburt, 1993) showed that descriptive experience sampling avoids the pitfalls that plagued turn-of-the-century introspection. Furthermore, I (Hurlburt, 1990) urged critics not to dismiss the descriptive

experience sampling method on the basis of informal attempts at replicating the procedure. Informal sampling attempts such as asking oneself on occasion, "What am I thinking right now?," are nearly always discouraging, leading the typical critic to believe that he or she would be unable to perform the sampling task. However, I reported that most participants found the actual task of responding to the random beep to be quite easy and unambiguous, stating that "unsuccessful [informal] attempts at thought sampling should not lead you to conclude that [descriptive experience] sampling . . . is impossible; but rather should lead you to an appreciation of the relative delicacy of the method" (Hurlburt, 1990, p. 269).

I now list some examples of the kinds of nomothetic conclusions that have been tentatively advanced by the descriptive sampling method. I (Hurlburt, 1990, 1993) showed that inner experiences sometimes have no symbols (words, images, feelings, etc.) at all, even though a thought content is clearly apprehended as ongoing at the moment. I called this phenomenon *unsymbolized thinking*—inner experiences that are clearly apprehended as being thoughts (e.g., "I'll wear my blue dress—no I'll wear the green one") but are not accompanied by the experience of words (e.g., *blue* or *dress*), images, feelings, or any other symbol (Hurlburt, 1993). The participant "just thinks" the content, with no other characteristics. I (Hurlburt, 1993) furthermore showed that the frequency of unsymbolized thinking tracks the depth of depression in some individuals whose depression is cyclic: When depression deepens, the frequency of unsymbolized thinking increases; when depression lessens, unsymbolized thinking decreases.

Many psychologists find the concept of unsymbolized thinking difficult to understand and more difficult to believe; many participants whose samples contain many unsymbolized thoughts believe prior to sampling that such thinking is impossible (Hurlburt, 1990, 1993). Part of this difficulty is the widely held prejudice that all thinking is in words. However, there is substantial evidence from conventional cognitive studies that some thinking is nonverbal (e.g., see Schooler & Melcher, 1995), but neither Schooler nor any other modern theorist posits the existence of unsymbolized thinking, which as yet is solely a descriptive-experience-sampling finding.

Doucette and Hurlburt (1993) sampled individuals with bulimia using the descriptive experience sampling method and found that their inner experiences were often multiple: There were frequently simultaneous occurrences of a few or many separable, identifiable inner happenings. As many as 5 or 10 separate, simultaneous thoughts were not uncommon. In some individuals with bulimia, self-induced vomiting was followed by a dramatic and immediate simplification of an inner experience.

Hurlburt, Happe', and Frith (1994) used the descriptive experience sampling method with a group of patients with Asperger syndrome—high-functioning individuals with autism. They found the striking result that if inner experience was reported at all, it involved only visual images (inner speech, feelings, etc. were never reported). These results have substantial implications for the study of autism, where considerable effort has been expended to ascertain whether it is possible for individuals with autism to understand the existence of mental events (e.g., Frith, 1991), let alone to describe them adequately. Furthermore, the conclusion that an inner experience is populated only by images offers a plausible explanation for the frequent observa-

tion that individuals with autism have extreme difficulty in taking another's point of view: It is extremely difficult, if not impossible, to represent another's point of view using only visual images.

As a final example, note that descriptive experience sampling calls the nature of hallucinations into question, at least for some patients. Visual hallucinations, for example, are generally thought to be visual experiences that are mistaken for reality. However, some of my (Hurlburt, 1990) patients with schizophrenia used the language of visual hallucinations to describe experiences that were not at all visual. For example, one patient with schizophrenia reported "seeing the Virgin Mary" (p. 215) at the moment of a beep. Casual interpretation of this report would be of a visual hallucination, but more careful probing revealed that this experience apparently had absolutely no visual characteristics at all. Actual visual hallucinations (experiences that do in fact involve visual phenomena) do occasionally occur at the moment of some beeps, but one must not assume that a person's use of visual terminology implies that that person experiences visual phenomena.

### Discussion

It has been shown that randomly sampling thoughts is a reliable and valid technique for exploring human experience and behavior and that despite the relatively small number of sampling studies (there are about 60 studies that analyze thought samples and another 60 ESM studies or so that sample experience but not thoughts), the method has proven itself to be capable of providing unique views of human nature, impossible to attain by other methods.

There are still many important issues to be resolved regarding the sampling method, not the least of which is whether or to what extent people can accurately report their momentary thought content and process. Did not psychology learn its lesson nearly a century ago that introspection is impossible, and was not that lesson reaffirmed more recently by the work of Nisbett and Wilson (e.g., 1977), among others? I (Hurlburt, 1990) argued that modern thought sampling is different from the earlier introspective studies because it asks participants simply to describe (quantitatively or descriptively) characteristics of thinking, whereas the earlier studies focused on theoretical inferences. I observed that most participants most of the time have no difficulty distinguishing among the various characteristics of inner experiences.

Is not the sampling process reactive, made so by its repeated-observation, in-the-natural-environment nature? A substantial number of reports observe that sampling is reactive in the sense that participants frequently find participation in sampling studies therapeutic (Hurlburt & Sippelle, 1978; Filstead, 1988; Figurski, 1992; Donner, 1992; Hurlburt, 1993). Fran's remarkable remission (described above) may have been due in part to the process of describing the details of her unpleasant inner experiences (Hurlburt, 1993). However, there is less evidence to decide the question of reactivity in the sense of the extent to which the sampling process may create, rather than simply record, the observed thoughts. I (Hurlburt, 1993) argued that descriptive experience sampling, the technique with the most experimenter-participant contact and therefore where the risk of reactivity seems greatest, is not excessively reactive, but clearly more

work needs to be done to decide the issue. Many sampling participants, to be sure, have reported a few thoughts that are directly related to the sampling experiment, but the number of such thoughts is usually extremely small, probably due to the random sequence and extended time periods over which sampling takes place.

It has been my intent to demonstrate that the sampling methods may provide unique perspectives on human experience. Many more sampling studies need to be performed, both to explore the adequacy of the sampling process itself and to use sampling to explore other content areas. It has not been my intent to claim that sampling is the best or the only method of examining thinking; an informed psychology can benefit from a variety of techniques. The benefits of the control and manipulation of variables and stimuli and the focus on external events that characterize experimental studies are well known and need not be belabored; what is at issue is whether those benefits have blinded psychologists to the countervailing benefits of ecological validity and the direct examination of inner experience.

One reason that sampling methods, despite their apparent nomothetic and idiographic value, have not been widely accepted by mainstream psychology has been the success of the behaviorist attack on introspection. Lashley (1923), one of the widely cited leaders of this attack, held that "introspection may make the preliminary survey, but it must be followed by the chain and transit of objective measurement" (p. 352). The critique of introspection by Lashley and many others has been so severe and so dominant that all introspection, including the "preliminary survey" that Lashley himself favored, has been banished from mainstream psychology. Even Lashley himself might now agree that allowing the experimental psychology "chain and transit" to lack introspective guidance may have been a major failing of the discipline and that it may be time to encourage experimental psychological studies to be guided by and validated against the ecologically valid sampling methods.

### References

- Beck, A. T., & Steer, R. A. (1990). *Beck Anxiety Inventory*. San Antonio, TX: The Psychological Corporation.
- Brunswick, E. (Ed.). (1952). The conceptual framework of psychology. *International encyclopedia of unified science* (Vol. 1). Chicago: University of Chicago Press.
- Csikszentmihalyi, M., & Figurski, T. J. (1982). Self-awareness and aversive experience in everyday life. *Journal of Personality*, 50, 15-28.
- Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the experience-sampling method. *Journal of Nervous and Mental Disease*, 175, 526-536.
- Delespaul, P. A. E. G. (1995). *Assessing schizophrenia in daily life: The experience sampling method*. Maastricht, The Netherlands: Universitaire Pers Maastricht.
- Delespaul, P. A. E. G., Reis, H. T., & de Vries, M. W. (1995). ESM time and RIR event sampling in social interactions. In P. A. E. G. Delespaul (Ed.), *Assessing schizophrenia in daily life: The experience sampling method* (pp. 116-132). Maastricht, The Netherlands: Universitaire Pers Maastricht.
- de Vries, M. W., & Delespaul, P. A. E. G. (1992). Variability of schizophrenia symptoms. In M. W. de Vries (Ed.), *The experience of psychopathology: Investigating mental disorders in their natural settings* (pp. 97-109). Cambridge, England: Cambridge University Press.
- de Vries, M. W., Delespaul, P. A. E. G., Dijkman, C. I. M., & Theunissen, J. (1986). Advances in understanding temporal and setting aspects



- of schizophrenic disorders. In F. Massimini & P. Inghilleri (Eds.), *L'esperienza quotidiana* (pp. 477–493). Milan, Italy: Franco Angeli.
- de Vries, M. W., Dijkman-Caes, C. I. M., Delespaul, P. A. E. G. (1990). The sampling of experience: A method of measuring the co-occurrence of anxiety and depression in daily life. In J. D. Maser & C. R. Cloninger (Eds.), *Comorbidity of anxiety and mood disorders* (pp. 707–726). Washington, DC: American Psychiatric Press.
- Dijkman-Caes, C. I. M., & Delespaul, P. A. E. G. (1995). Comparing ESM with diary assessments. In P. A. E. G. Delespaul (Ed.), *Assessing schizophrenia in daily life: The experience sampling method* (pp. 110–115). Maastricht, The Netherlands: Universitaire Pers Maastricht.
- Donner, E. (1992). Expanding the experiential parameters of cognitive therapy. In M. W. de Vries (Ed.), *The experience of psychopathology: Investigating mental disorders in their natural settings* (pp. 260–269). Cambridge, England: Cambridge University Press.
- Doucette, S., & Hurlburt, R. T. (1993). Inner experience in bulimia. In R. T. Hurlburt, *Inner experience in disturbed affect* (pp. 153–163). New York: Plenum Press.
- Figurski, T. J. (1992). Everyday self-awareness: Implications for self-esteem, depression, and resistance to therapy. In M. W. de Vries (Ed.), *The experience of psychopathology: Investigating mental disorders in their natural settings* (pp. 304–313). Cambridge, England: Cambridge University Press.
- Filstead, W. J. (1988). Monitoring the process of recovery using electronic pagers as a treatment intervention. In M. Galanter (Ed.), *Recent developments in alcoholism* (Vol. 6, pp. 181–191). New York: Plenum Press.
- Freeman, M., Csikszentmihalyi, M., & Larson, R. (1986). Adolescence and its recollection: Toward an interpretive model of development. *Merrill Palmer Quarterly*, 32, 167–185.
- Frith, U. (Ed.). (1991). *Autism and Asperger syndrome*. Cambridge, England: Cambridge University Press.
- Hnatiuk, S. H. (1991). Experience sampling with elderly persons: An exploration of the method. *International Journal of Aging and Human Development*, 33, 45–64.
- Hormuth, S. E. (1986). The sampling of experience in situ. *Journal of Personality*, 54, 262–293.
- Hunt, D. A., & Rosen, J. C. (1981). Thoughts about food by obese and nonobese individuals. *Cognitive Therapy and Research*, 5, 317–322.
- Hurlburt, R. T. (1976). *Self-observation and self-control*. Unpublished doctoral dissertation, University of South Dakota.
- Hurlburt, R. T. (1979). Random sampling of cognitions and behavior. *Journal of Research in Personality*, 13, 103–111.
- Hurlburt, R. T. (1990). *Sampling normal and schizophrenic inner experience*. New York: Plenum Press.
- Hurlburt, R. T. (1993). *Sampling inner experience in disturbed affect*. New York: Plenum Press.
- Hurlburt, R. T., Happe', F., & Frith, U. (1994). Sampling the form of inner experience in three adults with Asperger syndrome. *Psychological Medicine*, 24, 385–395.
- Hurlburt, R. T., & Melancon, S. M. (1987). How are questionnaire data similar to, and different from, thought-sampling data? Five studies manipulating retrospectiveness, single-moment focus, and indeterminacy. *Cognitive Therapy and Research*, 11, 681–703.
- Hurlburt, R. T., & Sippelle, C. N. (1978). Random sampling of cognitions in alleviating anxiety attacks. *Cognitive Therapy and Research*, 2, 165–169.
- Ingram, R. E. (1990). Self-focused attention in clinical disorders: Review and a conceptual model. *Psychological Bulletin*, 107, 156–176.
- Johnson, C., & Larson, R. (1982). Bulimia: An analysis of moods and behavior. *Psychosomatic Medicine*, 44, 341–351.
- Klinger, E. (1978–1979). Dimensions of thought and imagery in normal waking states. *Journal of Altered States of Consciousness*, 4, 97–113.
- Klinger, E. (1984). A consciousness-sampling analysis of test anxiety and performance. *Journal of Personality and Social Psychology*, 47, 1376–1390.
- Klinger, E., & Cox, W. M. (1987–1988). Dimensions of thought flow in everyday life. *Imagination, Cognition and Personality*, 7, 105–128.
- Kraan, H., Meertens, H., Hilwig, M., Volovics, L., Dijkman-Caes, C. I. M., & Portegijs, P. (1992). Selecting measures, diagnostic validity and scaling in the study of depression. In M. W. de Vries (Ed.), *The experience of psychopathology: Investigating mental disorders in their natural settings* (pp. 324–338). Cambridge, England: Cambridge University Press.
- Kroll-Mensing, D. (1992). *Differentiating anxiety and depression: An experience sampling analysis*. Unpublished doctoral dissertation, University of Minnesota, Twin Cities.
- Lashley, K. S. (1923). The behavioristic interpretation of consciousness II. *Psychological Review*, 30, 329–353.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we know: Verbal reports on mental processes. *Psychological Review*, 84, 231–259.
- Popper, K. (1963). *Conjectures and refutations*. New York: Basic Books.
- Prescott, S., Csikszentmihalyi, M., & Graef, R. (1981). Environmental effects on cognitive and affective states: The experiential time sampling approach. *Social Behavior and Personality*, 9, 23–32.
- Saltman, S. (1983). *Random sampling of cognitions in everyday living: A case study*. Unpublished master's thesis, University of Nevada, Las Vegas.
- Schooler, J., & Melcher, J. (1995). The ineffability of insight. In S. M. Smith, T. B. Ward, & R. A. Finke (Eds.), *The creative cognition approach* (pp. 97–133). Cambridge, MA: MIT Press.
- Shiffman, S., Fischer, L. A., Paty, J. A., Guys, M., Hickcox, M., & Kassel, J. D. (1994). Drinking and smoking: A field study of their association. *Behavioral Medicine*, 16, 203–209.
- van der Poel, E. G. T., & Delespaul, P. A. E. G. (1992). The applicability of ESM in personalized rehabilitation. In M. W. de Vries (Ed.), *The experience of psychopathology: Investigating mental disorders in their natural settings* (pp. 290–303). Cambridge, England: Cambridge University Press.

Received January 3, 1997

Revision received March 11, 1997

Accepted March 28, 1997 ■