Russell T. Hurlburt and Christopher L. Heavey

To Beep or Not To Beep

Obtaining Accurate Reports About Awareness

Abstract: *We begin by accepting that introspective evidence is important to cognitive science. However, as its history shows, introspection is risky, so methods should be used that minimize those risks. We argue that there are 13 ways that a beeper can reduce those risks, dividing those ways into three categories: time sampling* per se*, minimizing the reactive disturbance of evanescent phenomena, and aiding phenomenological fidelity. We turn aside six criticisms of beeper-based research, and describe five characteristics of a good beep.*

Readers of this special issue will accept that awareness is worthy of scientific study, if for no other reason than because awareness has long been held to be central to the scientific study of consciousness. For example, John Dewey (1906) wrote that 'conscious means aware; consciousness, the state of being aware' (p. 40). Natsoulas (1978) discussed six meanings of the term 'consciousness' and concluded that of these meanings, awareness is central and 'eminently worthy of our attention, perhaps as much as the nature of matter is worthy of the attention of other scientists' (p. 910). More recently, Searle (1990) meant by consciousness the subjective state of awareness, and Crick (1994) wrote that he 'used the terms awareness and consciousness more or less interchangeably' (p. 10, italics in original).

Thus we accept that awareness is important, but how do we gain access to it? The answer seems obvious: We need to introspect.

Introspection: Navigating Between the Scylla of 'Impossible' and the Charybdis of 'Easy'

It is well known that introspection has been banished for nearly a century from psychological investigation, exemplified by Nisbett and Wilson's scathing criticism, 'The accuracy of subjective reports is so poor as to suggest than any

Correspondence:
Russell T. Hurlburt, Dept. of Psychology, University of Nevada, Las Vegas, 4505 S. Maryland Parkway, Las Vegas, NV 89154-5030, USA. *Email: russ@unlv.nevada.edu*

Journal of Consciousness Studies, **11**, No. 7–8, 2004, pp. 113–28

introspective access that may exist is not sufficient to produce generally correct or reliable reports' (1977, p. 233). For the last decade or so, however, psychology has returned to introspective methods and consideration of consciousness and inner experience (Brewer, 1988; Cohen & Schooler, 1997; Dennett, 1991; Ericsson & Simon, 1980; 1984/1993; Searle, 1990; Marcel & Bisiach, 1988). Baars (1997) observed recently that 'the scientific race for consciousness is now on' (p. viii). It is as if the behaviourist dictatorship that has harshly suppressed studies of consciousness has finally been overthrown. Calls for introspection now abound, as the present volume attests.

We have argued elsewhere (Hurlburt & Heavey, 2001) that there are two mistakes to be made with respect to introspection: (1) believing that introspection is impossible, and (2) believing that introspection is easy. The radical exclusion of introspection from scientific method throws the baby out with the bath water. Thus we must disagree with Nisbett and Wilson's overly general conclusion cited above (although we noted in our 2001 paper that Nisbett and Wilson themselves carved out exceptions from their own sweeping conclusion). But the acceptance of introspective reports without adequate concern about whether they are accurate is just as mistaken, so we also must disagree with Chalmers, who seemed to imply that we can simply observe consciousness: 'We know about our own detailed and specific conscious experiences . . . we each have access to a rich source of data in our own case' (1996, p. 215–6). Although it is certainly true that each of us has access to rich data about our own inner experience, as scientists we should recognize that the lack of rigour used in 'gathering' these data leaves open the possibility of significant error. In fact, our own explorations have shown that people, including sophisticated psychologists and philosophers, are often substantially mistaken in their 'armchair' characterizations of their inner experience and quite surprised by what they observe when they employ a careful and systematic method for observing their inner experience.

Further evidence in support of the proposition that obtaining accurate reports of awareness is not trivially easy can be found in the many historical disagreements about the nature of awareness. For example, the debate over whether imaginal indeterminacy was possible continued for centuries, as summarized by Schwitzgebel (2002):

> An earlier debate, more familiar to most philosophers, also apparently reflecting fundamental disagreement about the experience of imagery, was the debate between Locke and Berkeley about abstract ideas. Locke seems to have felt that he could form an image of a triangle that is 'neither oblique, nor rectangle, neither equilateral, equicrural, nor scalenon; but all and none of these at once' (1689/1975, p. 596). Berkeley denied that he had the capacity to generate any such images (1710/1965). While Berkeley leaves open the possibility that Locke's imagery is just very different from his own, it is evident that Berkeley felt the more plausible view to be that his own failure to experience abstract images was universal and that Locke simply got it wrong about his own conscious experience (pp. 37–8).

Thus either Berkeley or Locke got it wrong. Because both were intelligent people trying quite hard to get it right, it follows that observers engaged in

armchair observation, even very sophisticated observers, can make substantial mistakes when they draw conclusions about inner experience.

Certainly it is not new to criticize armchair introspection, but despite frequent criticism, armchair observation remains common, and the faith in it remains unjustifiably strong. For example, there is a 'new' tendency to encourage scientists to use retrospective reports alongside brain imaging or other physiological measures (e.g., Jack & Roepstorff, 2002). But unless modern researchers can demonstrate that they are more perspicacious than Berkeley or Locke and others who have foundered in similar waters, we should be quite sceptical about their armchair introspections.

Then how should science navigate between this Scylla that tells us introspection is impossible and the Charybdis that lures us into thinking it is easy? Science should develop systematic methods that seek to improve the quality of introspective observation. This paper argues that random probes (usually 'beeps') of awareness may be useful in navigating the treacherous waters between the 'it's easy' and 'it's impossible' perspectives regarding obtaining accurate reports of awareness.

The Beeper as Navigational Aid

We have discussed the perils of casual observation of inner experience, and we now turn to discuss how a beeper can help to avoid many of these perils. At the outset, we wish to acknowledge that there is nothing either magic or unique about a beeper, including the beepers that we employ in our own beeper-based research (e.g. Hurlburt, 1990; 1993; 1997; Hurlburt & Heavey, 2002). Using a beeper does not guarantee making an accurate observation any more than a sextant or GPS guarantees that you won't run your boat aground. In fact, we think accurate observing requires a high level of skill regardless of whether a beeper is employed. Nonetheless we believe that a beeper is a valuable tool in the journey toward understanding awareness. We divide the potential benefits of using a beeper into three broad categories: time sampling *per se*, minimizing the reactive disturbance of evanescent phenomena, and phenomenological fidelity. Some of the categories have been discussed previously, for example by Stone and Shiffman (1994), Stone *et al*. (1991), Hormuth (1986), Larson and Csikszentmihalyi (1983), Barrett and Barrett (2001), Delespaul (1992), and Hurlburt (1990; 1993). But all these accounts are incomplete, and none focus on the beep as a tool for facilitating phenomenological fidelity.

Time sampling per se

The first two ways in which a beeper can provide navigational aids between supposing that introspection is impossible and that it is easy are the well known features of time sampling.

Aid 1: The beep can provide time samples. Time sampling is not new. Goodenough defined time sampling 75 years ago as 'the observation of everyday behaviour of an individual or a group of individuals for definite short periods of

time and the recording of the occurrence or non-occurrence of certain specified and objectively defined forms of behaviour during each of these periods' (1928, p. 23). Time sampling of inner experience is useful for the same reasons that time sampling is useful in time-and-motion studies, and need not be further described here. The first use of the time sampling of inner experience was reported by Locke and Jensen (1974), who used a freon-powered horn to interrupt physical education classes once per day. Students were instructed to write down their thoughts on conveniently placed clipboards. The first use of repeated, random or quasi-random sampling of inner experience began in 1974 independently by Klinger (1978–9) and by Hurlburt (1976; Hurlburt & Sipprelle, 1978).

Aid 2: The beep can come in the subject's own natural environment. A cornerstone of time sampling as practiced throughout behavioural science is that the samples should be taken in the environments that the subject naturally inhabits, rather than in experimental or otherwise artificial situations. Hormuth (1986) credits Brunswik (1949) with coining the term 'ecological validity' to describe this aspect of sampling. The Experience Sampling Method (ESM), as practiced by Csikszentmihalyi and his colleagues (e.g., Larson & Csikszentmihalyi, 1983) is the largest-scale set of ecologically valid explorations of inner experience.

Altmann (1971) was an early critic of laboratory studies, arguing that relying heavily on laboratory studies risked an inordinate sacrifice of generality and relevance. Pope (1978) showed that longer sequences of thought, and thought content more remote from the participant's immediate circumstances, were obtained when subjects were reclining rather than walking freely, thus demonstrating the importance of ecological validity: if variables such as posture affect thought characteristics, what other variables might have at least as large an effect?

Minimizing the reactive disturbance of evanescent phenomena

Many characteristics that are examined by time-sampling methods are robust, relatively long duration characteristics that are not likely to be immediately disturbed by the sampling method. For example, ESM asks subjects to report whether they are alone or not alone at the time of the beep. By contrast, the characteristics of primary interest to the readers of this volume are the evanescent, transient phenomena of thinking, feeling, sensing. The next five features of the beep can aid in the accurate capture of evanescent phenomena.

Aid 3: The beep can be a quick external signal. In this respect the beep can be like a flashbulb that slices through the ongoing stream of time at a precise moment. A method that allows the focus on particular moments has the potential to catch aspects of awareness that may be fleeting and of brief duration.

James (1890/1952) argued that attempts to capture awareness are difficult or impossible:

> As a snowflake crystal caught in the warm hand is no longer a crystal but a drop, so, instead of catching the feeling of relation moving to its term, we find we have caught some substantive thing, usually the last word we were producing, statically taken, and with its function, tendency, and particular meaning in the sentence quite

> evaporated. The attempt at introspective analysis in these cases is in fact like . . . trying to turn up the gas quickly enough to see how the darkness looks (p. 158).

Although James was right to recognize the difficulty of the introspective task, we now recognize that a flashbulb can in fact reveal at least some of what is present in the darkness (enlarged pupils, nocturnal animals, and so on), even if it cannot capture the darkness itself. Similarly a beep that marks a precise moment in an ongoing stream of awareness has the potential to reveal aspects of awareness that were previously unrecognised because, like the snowflake caught in the warm hand, these aspects existed only for a fleeting moment.

Aid 4: The beep can come at times that are independent of the subject's internal or external states. In armchair observation, you ask yourself, 'What am I thinking?' only at particular times, namely whenever motivational and situational conditions are ripe for the formation of that question. Perhaps, for example, you observe only at times where you are in a heightened readiness to observe. If so, then armchair observation will make it seem that you are always in heightened readiness. By contrast, a random or quasi-random beep comes at times unrelated to current motivation and therefore triggers the examination of a wider range of phenomena.

Aid 5: The beep can ask you to pay attention to a contemporaneous (or at least just immediately past) event, thus maximizing the chances that the fullest possible account of the event is still available to awareness. As John S. Mill observed more than 100 years ago, the capturing of immediately past events may be the best procedure for gaining access to awareness: 'A fact may be studied through the medium of memory, not at the very moment of our perceiving it, but the moment after: and this is really the mode in which our best knowledge of our intellectual acts is generally acquired. We reflect on what we have been doing when the act is past, but when its impression in the memory is still fresh' (Mill, 1882/1961, p. 64).

This assertion is consistent with our current understanding of memory processes, which indicates that the more time that intervenes between an event and its recollection, the more likely information will be lost or distorted due to failure to encode in long-term memory, decay, interference, or memory construction processes (e.g., Ericsson & Simon, 1980; 1984/1993; Kahneman & Tversky, 1982; Schwarz, 1990). Consistent with this, Hurlburt and Melancon (1987) showed in a beeper study that longer-term retrospective reports may contain distortions of frequency of categories of thinking and clarity of inner experience. Similarly Brewer (1988) showed that autobiographical memory decays over time, with ability to recall thoughts decaying fastest.

Schooler, Reichle, and Halpern (in press) studied 'zoning out' while reading. They asked participants in an experimental situation to read a passage and press a key whenever they zoned out — lapsed into daydreaming — while reading. They found using a beeper-type probe that people are likely to overlook events such as zoning out even though they have been specifically primed to pay attention to precisely those events. Thus we should have limited confidence in the ability of

people to recall accurately moments of awareness as they pass by in the stream of consciousness.

Aid 6: The beep can reduce the reliance on and disturbance by meta-cognition. Self-observation generally begins with a self-directed question such as, 'What am I thinking?' But the 'What am I thinking?' question *is* the thought process at that time, so the attempt to observe inner experience either results in the constant observation of the 'What am I thinking?' question itself, or it requires the difficult if not impossible task of setting that question aside and observing the remainder of the ongoing stream. By contrast, the beep can come to signify 'What am I thinking?' in a particularly compact form, and, because it comes from an external source, the question itself does not occupy a central aspect of the subject's inner activity. Of course the beep does disturb the ongoing experience to some extent (it has to be heard and processed), but that disturbance may be less than in other ways of exploring awareness.

Aid 7: The beep can precisely designate the time under examination. Shortly after Michelle Kwan fell during the 2002 Olympic figure skating, a television reporter asked her, 'What were you thinking about when you fell during that triple?' Ms. Kwan's answer to the reporter was, 'I was thinking, "Well, that's skating!"' Here we note that, like most armchair observers, the reporter didn't specify precisely which moment she was inquiring about: immediately prior to the fall (in which case the thinking might be the distraction that caused the fall); as the jump was actually happening (in which case the thinking might be about whether to abort the triple into a double); as she hit the ice (perhaps 'Oh, pshaw!'); immediately thereafter (perhaps angry at herself or her coach); and so on. Lack of precision in defining 'when' must therefore lead to lack of precision in the 'what' as well. The beep has the potential of being substantially more explicit about the exact time under consideration.

Phenomenological fidelity

Suppose that your task is to discover the characteristics of inner experience — simply to describe the phenomena that occur. There are six ways that the beep provides assistance in reporting accurately about phenomena. Descriptive Experience Sampling (DES; Hurlburt, 1990; 1993; 1997) is a method developed to obtain accurate reports of inner experience. As far as we know it is the only method using beepers that incorporates these six phenomenological fidelity characteristics of the beep.

DES involves giving participants a random beeper and training them to pay attention to their ongoing awareness at the moment of onset of the beep. DES participants are asked to record the characteristics of the awareness that was ongoing at the moment of the beep, typically by writing notes on a pad. Participants are then interviewed with the goal of developing high-fidelity accounts of their ongoing awareness. This sampling-then-interviewing process is repeated over a series of days until some sense of the nature of the individual's unique inner experience is obtained. Based on the findings of DES, Hurlburt and

Heavey (2002) developed a codebook of categories of the form of inner experience that occurred frequently across individuals. They showed that they could reliably categorize the most common forms of inner experience. Hurlburt, Koch, and Heavey (2002) showed that there is a connection between inner experience and external behaviour, and Hurlburt (1990; 1993) reported on the salient characteristics of individuals with noteworthy commonalities (e.g., schizophrenic, depressed, etc.).

It should be noted that as DES is focused on describing ongoing awareness, it is fundamentally different from the classic introspection performed by Titchener and others a century ago. For example, Schwitzgebel (2004 in this volume) presents in our view a fair and accurate account of Titchener's introspective efforts, thus providing the opportunity to compare and contrast DES with Titchenerian introspection.

Whereas both DES and Titchenerian introspection attempt to elicit accurate reports about experience, the two methods are substantially different in their targets. Titchener was primarily interested in what we might call the psychophysical aspects of consciousness, or the fundamental mental processes. Schwitzgebel's paper illustrates this; the examples cited are: comparing the relative brightness of two different colours; discerning a very low tone sensation from a sensation of atonal noise; making the quantitative assessment that two sensations are each an equal distance, in different directions, from a third; distinguishing difference (or combination) tones; reporting the characteristics of the 'flight of colours' (complex afterimages); and perceiving non-obvious visual illusions. That is, Titchener was primarily interested in discerning obscure fundamental processes.

By contrast, DES is *not* interested in discerning obscure fundamental processes, but instead is interested in reporting obvious incidental occurrences. If inner speech happens to be present, DES describes that speech; if an image happens to be present, DES describes that image; if a feeling happens to be present to awareness, then DES describes that feeling; and so on. DES does *not* attempt to explore any fundamental processes (perceptual acuity, etc.) that underlie or perhaps create the incidental occurrences caught by the DES beeper. DES is *not* interested in the obscure or the hard to detect. It is interested only in the obvious, the easily apprehendable; that is, it tries to obtain accurate reports about the incidental occurrences themselves.

Thus there is substantial contrast between Titchener's attempts to discern obscure fundamental processes and DES attempts to report obvious incidental occurrences. Titchener had to train his introspectors to *suppress* the very things that DES seeks to discover — whatever happens to be passing through awareness at any given moment. Titchener trained subjects to suspend those incidental awarenesses to be better able to focus on the fundamental processes behind them. DES trains subjects to ignore any interest in fundamental processes and simply pay attention to the details of whatever incidental occurrence happens to be ongoing. The DES task is substantially easier than that of Titchener's introspection, and this accounts for the large differences in training time required. We

believe that most subjects can be adequately trained in the DES method in a few sampling days.

Just because the DES targets are obvious doesn't imply that giving accurate reports about them is trivially easy. Just as did Titchener's introspectors, DES subjects have to learn to avoid the 'stimulus error', mistaking a characteristic of experience for a feature of the environment. For example, if a DES subject reports seeing an inner image of his living room, he has to learn to avoid letting what he knows about his actual living room influence what he says he sees in the image. Furthermore, DES subjects have to learn to be consistent in the referents of words they use to describe their inner experiences, and DES interviewers have to learn what those referents actually are. For example, DES subjects use the word 'thinking' in a variety of ways; for one subject, it means 'inner speech'; for another, it means 'visual images'; for a third it means 'emotional experience' (Hurlburt & Heavey, 2001).

It is a fact of the human condition that what is obvious at the moment is frequently (for whatever reason) systematically overlooked by retrospection. The DES explorations are perhaps most useful in those situations where the once-immediately-obvious has become the now-retrospectively-forgotten. But that does not change the fact that DES limits itself to exploring the immediately obvious. In so doing, it avoids some of the major pitfalls that plagued Titchener's introspection.

Aid 8: The beep can assist in setting aside preconceptions by asking participants to pay attention to whatever happens to be ongoing at the moment of the beep. By contrast, many if not most formal and informal psychological investigations are in search of something relatively specific, something preconceptually believed to occur. Although there is parsimony in this strategy, there is also great risk of missing what you do not know to look for or substantially disturbing or altering the nature of the awareness you do find. For example, we discussed above the zoning-out-while-reading study by Schooler, Reichle, and Halpern (in press). That study is exemplary in its use of a beep to explore immediately occurring phenomena. Here, however, we observe that that study began with powerful preconceptions that cannot be dispelled within the design of that study. For example, Schooler, Reichle and Halpern believed at the outset that 'when one is zoning out, they are experientially conscious of whatever topic has grabbed their attention, while at the same time lacking meta-consciousness of the fact that they are zoning out. At some point during the reading episode, they suddenly become meta-conscious that they have been zoning out and realize that for some time they have been reading without comprehension' (Schooler *et al*., in press, quoted by permission). We believe that such meta-conscious awareness is by no means necessary: it is quite possible that a person could read, think about something unrelated to reading, and then return to reading *with no intermediate meta-cognitive awareness*. Whether that happens, or how often that happens, is a central issue in a study of meta-cognition, but the Schooler, Reichle, and Halpern methodology makes it impossible to observe meta-cognition as it naturally occurs or doesn't occur. They explicitly instructed their subjects to take

a meta-cognitive stance: subjects were to press a key every time they zoned out. In essence, they instructed their subjects to engage in precisely the behaviour that they later 'discovered'. The beep can provide an alternative that avoids that preconceptual trap. For example, have subjects read in the laboratory and beep them at random intervals. Their task is simply to describe their awareness at the precise moment of the beep, whatever that awareness happens to be. No mention of zoning out or any other phenomena need be made — the study seeks to discover the phenomena that occur while reading, whatever those phenomena are. If at a beep the subject is thinking about something unrelated to the reading, the subject will describe that experience. If not, not. If such unrelated thinking occurs frequently, then we would be justified in concluding that 'zoning out' is frequent. If at a beep the subject is thinking about his own mental processes, the subject will describe that experience. If not, not. If such self-observation occurs frequently, then we would be justified in concluding that meta-cognition is frequent and to explore its relatively undisturbed characteristics.

Certainly we do not wish to single out the Schooler, Reichle, and Halpern study as being exceptional in this regard. Many, perhaps nearly all, psychological investigations begin with preconceptions and then attempt to verify them. We do wish to point out that that is a risky, quite possibly extremely misleading, approach to scientific investigation, particularly for sensitive phenomena such as awareness. If our task is to understand consciousness, we must observe the phenomena as closely as possible to the way they naturally occur.

Aid 9: Repeated observations can lessen the importance of any single observation. Repeated observations are of course the cornerstone of time sampling (see Aid 1). But the repeated nature of beep-triggered observations also facilitates the setting aside of preconceptions by lessening the relative importance of any single observation. Suppose the beep occurs and at the moment there is some surprising phenomenon occurring — words seem to be being imagined backwards, for example. The repeated-observation nature of the sampling method can allow the investigator to take a dispassionate, unattached, uninvolved view of the phenomenon. 'That seemed like words coming out backward? We don't really have to worry too much about it. If backward words is an important phenomenon, it will reappear at a subsequent beep. If not, not.' The phenomenon at any single beep becomes important only if similar phenomena occur at other beeps.

Aid 10: The beep can allow an adequate time to examine and/or describe a particular moment. This is by obvious contrast to think-aloud methods, which presume that the stream of external speaking can 'keep up' with the important details of inner experience. However, DES studies (e.g., Hurlburt, 1993; 1997) show that inner experience is often much more complex. Multiple simultaneous visual images, for example, occasionally occur, and such experiences could not possibly be adequately characterized by think-aloud methods.

Aid 11: The beep can trigger observation, not manipulation. Most modern psychological studies involve experimental or situational manipulation. For example, Suchan and his colleagues (2002) used fMRI to study visual imagery:

> The subject's task consisted of drawing imagined lines connecting encircled numbers in ascending order and estimating the number of lines crossing. Compared with a control task in which there were no crossed lines, there were significant rCBF increases in the cingulate gyrus, the adjacent superior frontal gyrus and in the left inferior parietal cortex. The rCBF changes of the latter area correlated with task performance time. Since these activation areas are close to those in imagery of movement trajectories, the authors concluded that they appear to be a subsystem for processing mental visuospatial images (Suchan *et al*., 2002, p. 163).

Whereas it is certainly plausible to assume that the subjects in this study were engaged in a visual imaging task, there may well be other non-image ways to approach the line-crossing task. This study cannot determine the individual differences of subjects in this regard. Furthermore, imagining line crossings is probably a highly unusual activity for most subjects, quite different from their naturally occurring visual imaging. The beep could be used to collect more natural images as follows: trigger the fMRI data collection and fractions of a second later sound a beep asking the subject to describe his or her experience that was ongoing a fraction of a second earlier. Select for fMRI analysis those samples that happen to include visual imagery and discard the rest. To our knowledge, such an investigation has not been performed.

Aid 12: The beep can identify specific concrete moments on which to test preconceptual general notions. Armchair observation often is based on general notions, sometimes sweeping ('What is my thinking usually like?') and sometimes less sweeping but still general ('What was I thinking while I solved that problem?'). Whatever those general notions are, they are likely to be brought into play any time the armchair observer thinks about his or her inner experience. These types of preconceptions can be quite resistant to change and often exist in the face of considerable contrary evidence. The beeps however, can provide specific, concrete instances against which preconceptions can be repeatedly tested. Because the several moments encountered on the repeated occasions that the beep facilitates are all different from each other at least to some extent, they encourage the observer to pay close attention to those preconceptual general notions as they are highlighted by the differing perspectives of the several moments.

Aid 13: The repeated-occasion nature of the beep allows inductive generalizations. If careful examination of a beeped moment at time 1 reveals characteristic X, and that same characteristic X is also revealed at times 3, 5, 6, 8, and so on, then it is possible to form a truly inductive generalization such as, 'Characteristic X is a frequent occurrence for this subject.' Armchair observations frequently form the basis of statements that look like inductive generalizations but are actually more reflective of stubborn preconceptions than repeated direct observation.

Criticisms of Beeper-based Research

Although beeper-based strategies for exploring awareness have existed for decades, they are still relatively rare. Why? At least partially because of the

following criticisms. In an attempt to clarify the appropriate limits of beeper-based strategies, we address these criticisms. We use examples from our own DES studies.

Criticism 1: The beep surprises the subject, who is not in a state of readiness to observe. To observe phenomena accurately, it is desirable both (a) to have circumstances favourable for observation (subject immediately prepared and highly trained, known stimuli explicitly presented as in a laboratory, etc.); and (b) to have circumstances favourable for the phenomenon to emerge (subject uninfluenced by any preconceptions transmitted during training, subject in his or her natural environments and engaged in his or her own natural behaviours, etc.). It may be impossible to fully meet both requirements simultaneously. We do note that after a day or two of experience, most of our DES subjects report that it seems to them that they can immediately observe their experience when the beep occurs — that is, that they are not surprised or disturbed by the beep, and they are able to note the experience that was ongoing at the moment of onset of the beep. Such reports of course do not make it so, but it is true that moderately experienced DES subjects seem to be able to respond to the beep quite easily.

Criticism 2: The random nature of the beep makes it likely that infrequent but important characteristics will be missed. This is unquestionably true for some kinds of phenomena. However, many observers of DES have been surprised at how frequently the random darts of the beep actually hit some seemingly rare phenomenon of interest. Our understanding of this is that important but seemingly rare phenomena are actually more frequently occurring than was thought. Clearly more research is necessary on this point, but in the meantime it is wise not to rule out a sampling-based approach until it has been actually established that the phenomenon will not be found using sampling. Stone and Nicolson (1987) recommended combining sampling with continuous recording (diary-type methods) as the most useful strategy.

Criticism 3: Inner experience is a stream, but the beeper approach makes it appear like a series of moments. It is certainly true that most reports of DES subject make it appear that experience is saltatory, more like a series of beads on a string than a continuous stream. Whether that is an artefact of the method or whether that is the way experience is needs further clarification. It is certainly possible that for some, perhaps most people, awareness jumps from one experience to the next with little or nothing in between. It is also possible that DES is by its nature unable to observe the actual stream-like characteristics of awareness. Further investigation is necessary here.

Criticism 4: There is no way of knowing how much the beep disturbs inner experience. This criticism is undeniably true — the beep does in fact disturb inner experience. DES tries to minimize this disturbance as much as possible by focusing on the moment of awareness immediately prior to the occurrence of the beep (what we call the 'last undisturbed moment' prior to the beep). Although this requires a slight 'rewinding' of the memory trace of awareness, most participants report little difficulty in doing this, which is consistent with our understanding of short term memory (Atkinson & Schiffrin, 1968). Furthermore, there

are sometimes external sources against which the claim of validity of inner experience reports can be tested. For example, Hurlburt (1993, 1997) discussed the case of Fran, concluding that Fran had no figure/ground phenomenon in her inner (and outer) experience. Hurlburt (1997) discussed five 'idiographic validity' reasons that the claim of no figure/ground should be believed. Even if Hurlburt's (1997) claims are true, this does not definitively rule out the criticism that the beep disturbs experience. However, we believe that it does limit the extent of that criticism.

Criticism 5: Responding to the beep puts the subject in a reflective stance, and thus beep-triggered introspection does not explore unreflective consciousness. That is absolutely true, and that is a primary reason that we refer to the subject matter of DES investigations to be 'inner experience' rather than 'consciousness.' Theories of consciousness require an extrapolation from reflective observations of awareness; the question here is whether the beeper helps to build a solid foundation from which to extrapolate. It is important to keep this fair criticism separate from a frequent but unfair confound with long-term retrospection and armchair observation.

Criticism 6: Reports about inner experience, whether beeped or not, are private, and therefore unreliable. The extent to which one should credit reports about inner experience is of fundamental importance. Hurlburt and Heavey (2002) explored the reliability of DES by independently interviewing subjects about the same beeped moments. The single-sample interobserver-reliability kappas for the five most frequently occurring characteristics of inner experience ranged from .52 to .92 with a median of .76. Spearman–Brown adjustment showed that reliabilities for typical 19-sample averages would range from .92 to .98, comparable to highly reliable questionnaires.

Hurlburt and Heavey (2002) thus demonstrated that ratings of inner experience can be highly reliable, but they did not measure the extent to which the reports about inner experience actually reflected the contents of awareness. That is a limitation of sampling-based and all other means of exploring inner experience. The fundamental question thus remains: To what degree should we believe the report of a subject? There is always a degree of doubt, and conviction is therefore always short of 100%, whether the reports are about inner experience or outer experience. Why does New York physicist Mary credit Pennsylvania physicist William's voltmeter reading? It is not because the reading was public — in fact, neither Mary nor anyone else was present when William made the reading. Mary credits William because he demonstrates knowledge, care, and skill regarding voltmeters. When William says, 'I replaced the voltmeter wires with new ones and obtained the same reading,' or 'I used two different voltmeters and obtained the same reading,' Mary's conviction rises because she sees that William is demonstrating that he knows the frequent causes for voltmeter-reading failure, and has exercised appropriate care to rule out those possibilities.

Certainly included in Mary's crediting of William is the fact that voltmeter readings *can* be public: Mary might know that in certain specified situations (such as in university physics classes) William obtained the same voltmeter

readings as did a qualified instructor. That is an important consideration, but it does not override all other considerations.

The question then becomes, how does one demonstrate knowledge, care, and skill when giving reports about inner experience? There is no definitive answer to that question, any more than there is a definitive answer to the question about how to demonstrate knowledge, care, and skill when using voltmeters. But it seems reasonable that the demonstration of interrater reliability is a first step, along with demonstrated sensitivity to other issues surrounding sampling.

Characteristics of a Good Signal

Suppose that in a particular situation a researcher believes that the reasons to beep outweigh the criticisms. What should the characteristics of the beeping signal be? We do not believe that the signal must necessarily be an audible beep. We discuss now five desirable characteristics of a good signal.

Characteristic 1: Unambiguity. Whatever is used for the signal must be unambiguous in the sense that the signal must mean 'Pay attention to your awareness now!' and nothing else. Thus it is not adequate to use, for example, a telephone ring to cue self-observation. At the moment of the telephone ring, the subject must in some way decide *not* to engage in the usual telephone-ring behaviour and *instead* pay attention to inner experience. That decision process substantially complicates the attentional task.

Characteristic 2: Easy detectability. The subject must not have to struggle to discern whether the signal exists or not. For example, if the beeper is too soft, the subject must engage in a task that is something like, 'Is that the beep? Is that the beep? Yes that's the beep!' but by now the evanescent characteristics of the experience have disappeared. If the signal is too loud, a startle response can obliterate ongoing experience.

Characteristic 3: Rapid onset. As suggested above, it is desirable to define with as much precision as possible the exact moment under investigation. That requires the signal to have a clearly defined rapid onset. That makes the kind of vibrator typically used in pagers not desirable, for example. The subject's ability to discern precisely the onset of the vibration is not adequate.

Characteristic 4: Privacy. The subject's task at the moment of the beep is to pay attention to his or her ongoing awareness as undisturbed as possible. Thus that paying attention is a solitary, private task. If the signal occurs in a public way, for example, a beep audible to all people in the vicinity, then the response to the beep is complicated by the social implications: 'What will I tell George when he asks me what that beep caught?' DES solves this issue by delivering the signal through a transistor-radio-type earphone.

Characteristic 5: Portability. If the object is to understand experience as it naturally occurs, then the signal must be able to follow the subject easily to wherever that experience takes place.

Discussion

We wholeheartedly endorse the current resurgence of scientific interest in inner experience. Our own research has shown to our satisfaction that this is fertile ground. Thus for example we strongly support the intent behind calls for the collaboration between objective and introspective science, such as that by Jack and Roepstorff (2002):

> [Incorporating retrospective accounts alongside fMRI procedures] serves to draw attention to the subject's experience. Post-experimental interviews (which clearly rely on episodic recall, hence our use of the term 'retrospection') are not well suited to probing the finer details of experience (such as variations between trials), or to statistical analysis. However they can provide a 'crude' or broad picture of the experiential states involved. At the very least, this should encourage experimenters to think more carefully before they speculate about the relevance of their objective measures to 'awareness' or other subjective phenomena. At best, retrospective reports will considerably enrich experimenters' understanding of 'what it is like' to do the task, potentially revealing unexpected and important experiential phenomena. Experimenters can then frame detailed hypotheses regarding these phenomena, which can be tested using 'tailored' introspective methodologies (p. 335).

However, we must caution against complacently presuming that armchair observations of the kind that Jack and Roepstorff suggest 'provide a "crude" or broad picture of the experiential states involved'. While that may be true in some instances, it can be dramatically untrue in others. For example, DES has shown that a frequent characteristic of inner experience is 'unsymbolized thinking,' the engaging in a process clearly apprehended as a thought but that does not involve words, images, or any other symbols (Hurlburt, 1993; 1997). In fact, Heavey, Hurlburt, and Cavenagh (2002) have observed that approximately a quarter of all inner experiences include unsymbolized thinking. While there are exceptions, most armchair observers either systematically neglect to observe such phenomena or actively hold that such thinking is impossible. Perhaps most striking is the fact as revealed by DES that many subjects who themselves engage in unsymbolized thinking frequently *themselves believe prior to sampling that unsymbolized thinking is impossible*. If these subjects had engaged in nonsampling retrospection of the kind called for by Jack and Roepstorff, they would never have reported unsymbolized thinking. They would not have given a 'crude or broad' description of it, they would have mistakenly stated that it does not exist.

We have identified 13 ways that beeping can aid in the introspective enterprise. We wish to emphasize that that does *not* imply that anyone who uses a beeper has direct access to the truth about inner experience. In this respect, the beeper is like the laser as a surgical tool: the laser is quite effective, allowing surgeries to be performed that were impossible without it. But you still want it to be operated by a skilled practitioner. The exploration of inner phenomena is in fact difficult: events are private, evanescent, and sensitive; people (subjects and investigators) hold strong but untrue preconceptions about inner experience; and

so on. The beeper is a useful tool in the exploration toolbox, but it does not necessarily solve every problem, nor does it guarantee the fidelity of the results.

References

Altmann, J. (1971), 'Observational study of behaviour: Sampling methods', *Behaviour*, **49**, pp. 227–267.

Atkinson, R.C., & Schiffrin, R.M. (1968). 'Human memory: A control system and its control processes', in K. Spence (Ed.), *The Psychology of Learning and Motivation*, vol. 2, (New York: Academic Press).

Baars, B.J. (1997), *In the Theater of Consciousness* (New York: Oxford University Press).

Barrett, L.F. & Barrett, D.J. (2001), 'An introduction to computerized experience sampling in psychology', *Social Science Computer Review*, **19**, pp. 175–85.

Berkeley, G. (1710/1965), 'A treatise concerning the principles of human knowledge', in C.M. Turbayne (Ed.), *Principles, Dialogues, and Philosophical Correspondence* (New York: Macmillan).

Brewer, W.F. (1988), 'Memory for randomly sampled autobiographical events', in U. Neisser & E. Winograd (Eds.), *Remembering Reconsidered* (New York: Cambridge University Press).

Brunswik, E. (1949), *Systematic and Representative Design of Psychological Experiments* (Berkeley and Los Angeles: University of California Press).

Chalmers, D. (1996), *The Conscious Mind* (New York: Oxford University Press).

Cohen, J.D. & Schooler, J.W. (1997), *Scientific Approaches to Consciousness* (Mahweh, NJ: Erlbaum).

Crick, F. (1994), *The Astonishing Hypothesis* (New York: Scribner).

Delespaul, P.A.E.G. (1992), 'Technical note: Devices and time-sampling procedures', in M. deVries (Ed.), *The Experience of Psychopathology: Investigating Mental Disorders in their Natural Settings* (Cambridge: Cambridge University Press).

Dennett, D.C. (1991), *Consciousness Explained* (Boston, MA: Little, Brown).

Dewey, J. (1906), 'The terms "conscious" and "consciousness"', *Journal of Philosophy, Psychology, and Scientific Methods*, **3**, pp. 39–41.

Ericsson, K.A. & Simon, H.A. (1980), 'Verbal reports as data', *Psychological Review*, **87**, pp. 215–51.

Ericsson, K.A. & Simon, H.A. (1984/1993), *Protocol Analysis: Verbal Reports as Data* (Cambridge, MA: Bradford Books/MIT).

Goodenough, F.L. (1928), 'Measuring behaviour traits by use of repeated short samples', *Journal of Juvenile Research*, **12**, pp. 230–35.

Heavey, C.L., Hurlburt, R.T. & Cavenagh, N.A. (2002), 'An exploration of the form of unaltered inner experience', Poster presented at Western Psychological Association, Irvine, CA.

Hormuth, S.E. (1986), 'The sampling of experiences *in situ*', *Journal of Personality*, **54**, pp. 262–93.

Hurlburt, R.T. (1976), 'Self-observation and self-control', Unpublished doctoral dissertation, University of South Dakota.

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. (1997), 'Randomly sampling thinking in the natural environment', *Journal of Consulting and Clinical Psychology*, **65**, pp. 941–9.

Hurlburt, R.T. & Heavey, C.L. (2001), 'Telling what we know: Describing inner experience', *Trends in Cognitive Sciences*, **5**, pp. 400–403.

Hurlburt, R.T. & Heavey, C.L. (2002), 'Interobserver reliability of Descriptive Experience Sampling', *Cognitive Therapy and Research*, **26**, pp. 135–142.

Hurlburt, R.T., Koch, M. & Heavey, C.L. (2002), 'Descriptive Experience Sampling demonstrates the connection of thinking to externally observable behaviour', *Cognitive Therapy and Research*, **26**, pp. 117–134.

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**, pp. 681–703.

Hurlburt, R.T. & Sipprelle, C.N. (1978), 'Random sampling of cognitions in alleviating anxiety attacks', *Cognitive Therapy and Research*, **2**, pp. 165–9.
Jack, A.I. & Roepstorff, A. (2002), 'Introspection and cognitive brain mapping: From stimulus-response to script-report', *Trends in Cognitive Sciences*, **6**, pp. 333–9.
James, W. (1890/1952), *The Principles of Psychology* (Chicago: Encyclopedia Britannica).
Kahneman, D. & Tversky, A. (1982), 'The simulation heuristic', in D. Kahneman & A. Tversky (Eds.), *Judgment Under Uncertainty: Heuristics and Biases* (New York: Cambridge University Press).
Klinger, E. (1978–1979), 'Dimensions of thought and imagery in normal waking states', *Journal of Altered States of Consciousness*, **4**, pp. 97–113.
Larson, R. & Csikszentmihalyi, M. (1983), 'The experience sampling method', *New Directions for Methodology of Social and Behavioral Science*, **15**, pp. 41–56.
Locke, J. (1689/1975), *An Essay Concerning Human Understanding* (Oxford: Clarendon).
Locke, L.F. & Jensen, M.K. (1974), 'Thought sampling: A study of student attention through self-report', *The Research Quarterly*, **45**, pp. 263–275.
Marcel, A. J. & Bisiach, E. (ed. 1988), *The Concept of Consciousness in Contemporary Science* (New York: Oxford University Press).
Mill, J.S. (1882/1961), *Auguste Comte and Positivism* (Ann Arbor: University of Michigan)
Natsoulas, T. (1978), 'Consciousness', *American Psychologist*, **33**, pp. 906–14.
Nisbett, R.E., & Wilson, T.D. (1977), 'Telling more than we know: Verbal reports on mental processes', *Psychological Review*, **84**, pp. 231–59.
Pope, K.S. (1978), 'How gender, solitude and posture influence the stream of consciousness', in K.S. Pope & J.L. Singer (Eds.), *The Stream of Consciousness* (New York: Plenum)
Schooler, J.W., Reichle, E.D. & Halpern, D.V. (in press), 'Zoning-out during reading: Evidence for dissociations between experience and meta-consciousness', in D. Levin (Ed.), *Visual Meta-Cognition: Thinking About Seeing* (Westport, CT: Praeger)
Schwarz, N. (1990), 'Assessing frequency reports of mundane behaviors: Contributions of cognitive psychology to questionnaire construction', in C. Hendrick & M.S. Clark (Eds.), *Research Methods in Personality and Social Psychology* (Newbury Park, CA: Sage)
Schwitzgebel, E. (2002), 'How well do we know our own conscious experience? The case of visual imagery', *Journal of Consciousness Studies*, **9** (5–6), pp. 35–53.
Schwitzgebel, E. (2004), 'Introspective training apprehensively defended: Reflections on Titchener's lab manual', *Journal of Consciousness Studies*, **11** (7–8), pp. 58–76.
Searle, J. (1990), 'Who is computing with the brain?', *Behavioral and Brain Sciences*, **13**, pp. 632–642.
Stone, A.A. & Shiffman, S. (1994), 'Ecological momentary assessment (EMA) in behavioral medicine', *Annals of Behavioral Medicine*, **16**, pp. 199–202.
Stone, A.A., Kessler, R.C. & Haythornthwaite, J.A. (1991), 'Measuring daily events and experiences: Decisions for the researcher', *Journal of Personality*, **59**, pp. 575–607.
Stone, P.J., & Nicolson, N.A. (1987), 'Infrequently occurring activities and contexts in time use data', *The Journal of Nervous and Mental Disease*, **175**, pp. 519–25.
Suchan, B., Yagueez, L., Wunderlich, G., Canavan, A.G.M., Herzog, H., Tellmann, L., Hoemberg, V., Seitz, R.J. (2002), 'Neural correlates of visuospatial imagery', *Behavioural Brain Research*, **131**, pp. 163–8.