Sleep
12(5):458–474, Raven Press, Ltd., New York

The Detection of Sleep Onset: Behavioral, Physiological, and Subjective Convergence

Robert D. Ogilvie, Robert T. Wilkinson, and Susan Allison

Psychophysiology Section, Medical Research Council Applied Psychology Unit, Cambridge, U.K.

Summary: In this report, a sleep deprivation/multiple arousal paradigm was used in which response time (RT) and respiratory and electroencephalographic (EEG) measures were combined with a continuous behavioral index of arousal (a deadman switch) and frequently repeated Stanford Sleepiness Scale ratings to examine the process of falling asleep. Sleep was defined behaviorally as failure to respond to the faint auditory RT cue. Although response rates decreased significantly as EEG stages passed from W through 1 to 2, responding continued in both light "sleep" stages. Respiratory, subjective, and DM changes were more pronounced between Stages W and 1 than between Stages 1 and 2. If the criterion for wakefulness is cognitive response to external stimulation, accurate distinctions between sleep and wakefulness can only be made in EEG Stages 3, 4, and rapid eye movement sleep. If EEG is the criterion, then the data suggest that cognitive response is possible during Stages 1 and 2 "sleep". The concept of a Sleep Onset Period, characterized by lengthening response times and intermittent response failure (thereby reflecting neither true sleep nor wakefulness), may provide a useful resolution to this definitional dilemma. **Key Words:** Sleep—Sleep onset—Sleep onset period—S/W definition.

The transition from wakefulness to sleep has remained ambiguous and even controversial when defined solely in terms of electrophysiological measures. Davis et al. (1) maintained that precise identification of a sleep onset (SO) point was impossible since the process of sleep/wakefulness (S/W) was continuous. Although widely accepted as the standard method for detecting the change from wakefulness to sleep, the electroencephalogram (EEG)-based criteria of alpha reduction proposed by Dement and Kleitman (2) and Rechtschaffen and Kales (3) has not gone entirely without criticism. For example, Agnew and Webb (4,5) argued that the onset of Stage 2 was a better indicator of SO; similarly, Johnson (6) used the presence of spindles to define sleep. However, since neither the diminution of alpha (3) nor the appearance of spindles and K-complexes (6) has been adequately assessed as an SO index, and given the continu-

Accepted for publication March 1989.
Address correspondence and reprint requests to Dr. R. D. Ogilvie at Psychology Department, Brock University, St. Catharines, Ontario, Canada L2S 3A1.

ing disagreement as to which (if any) EEG criterion can be used to identify SO accurately, it is surprising that behavioral corroborations of EEG-based definitions of SO have been virtually nonexistent.

Recently, more accurate detection of SO has been accomplished by assessing behavioral measures in addition to standard EEG recordings (7). Indeed, these investigations have shown that a behavioral measure of arousal, specifically, discrete responses to frequently presented faint tones (5–10 dB above background), can successfully serve as an indicator of SO and therefore replace or complement the somewhat vague EEG-based description of the transition from wakefulness to sleep. Conclusions from these investigations have led to the position that the process of SO is best described as a sequence of interrelated events. Ogilvie (24) argued that the transition between relaxed, drowsy wakefulness and unresponsive sleep be defined as a Sleep Onset Period (SOP).

Differences in respiratory rate have also been demonstrated between wakefulness and sleep and, moreover, changes in respiratory patterns have proven useful in the early detection of SO. Timmons et al. (8) identified three phases of respiratory activity that were related to three EEG phases. They reported that abdominal and thoracic amplitudes were very comparable while subjects were in drowsy Stage 1 EEG, but late in Stage 1, and definitely by the time Stage 2 sleep had begun, thoracic amplitudes exceeded abdominal levels in 10 of their 11 subjects.

Additional refinements were made by Naifeh and Kamiya (9,10); they added measurement of alveolar CO_2 tension to the EEG and strain gauge system of Timmons et al.(8). Naifeh and Kamiya (9) concluded that "that the process of sleep onset is intrinsically associated with central neural mechanisms controlling respiration" (p. 71).

Although there are consistent indications in the aforementioned literature that fundamental changes in respiration parallel changes in EEG-assessed arousal, the present research is the first we know of to assess both systems during SO by independent behavioral and subjective validators of sleep and wakefulness. Ogilvie and Wilkinson (7) showed that decreases in abdominal (A) amplitude relative to thoracic (T) respiratory activity and performance lapses are generally, but not exclusively, associated with the development of EEG Stage 1 sleep. They too found that the assessment of this respiratory measure helps to identify SO more accurately. Furthermore, by integrating behavioral, respiratory, and standard EEG measures (7), a more comprehensive and refined definition of SO was advanced. Since subjective assessments of SO have received considerable attention recently (10,11), and have added to the understanding of the process of falling asleep, including this dimension along with the behavioral and physiological indicators may still further advance the accurate detection of SO. Such an integrated approach is undertaken for the first time in the present study in the hope that the convergence of behavioral, physiological, and subjective assessments of SO will broaden understanding of the transition from wakefulness to sleep and thereby increase our ability to define and predict SO with a reasonable degree of confidence.

In the present investigation, the Stanford Sleepiness Scale (SSS) (12) was used to assess sleepiness during the process of falling asleep. Modified so that ratings could be given verbally until sleep rendered the subject silent, the SSS was used for the first time here in conjunction with other behavioral and electrophysiological measures of SO, and a continuous measure of wakefulness was added to increase the behavioral assessment of SO. The focus of the present investigation was also on the behavioral response definition of sleep and on contrasting that definition with EEG-based S/W distinctions.

Also, the experimental design was such that the intercorrelations could be calculated for experiential, objective behavioral, and physiological variables. The general intention was similar to Kamiya's (13) hope ". . . that the alignment of phenomena from each of the three realms—behavioral, experiential, and physiological—will stimulate their eventual theoretical integration" (p. 147).

METHODS

Subjects

Nine female and three male volunteers participated in this investigation. Four were recruited as a result of a letter mailed to members of the MRC APU subject panel. Of the other eight, seven were college students and one was a former MRC employee. The subjects ranged in age from 20 to 46 years ($\bar{x}$ = 26.9, SD = 9.1).

Two additional subjects completed the experimentation, but data from one were excluded from analyses because of apparatus failure, while scores from the other were omitted because the experimenters could not be confident that the volunteer had attempted to follow instructions during the testing sessions.

Apparatus

A 16-channel SLE electroencephalograph (model 180) and a 7-channel Ampex AM/FM tape recorder (model S300) were the major recording instruments. A Gould dual beam oscilloscope was used to monitor all signals before and during testing. The sleep chamber located inside the main EEG recording room measured 1.3 × 2.5 m.

The SSS, developed by Hoddes et al. (12) and validated on sleep-deprived subjects (14), is a simple Likert 7-point rating scale designed to assess sleepiness. Used orally in the present study, the volunteer said a number from one to seven that represented his sleepiness at that moment on the clearly defined steps of the scale.

Videotape equipment permitted constant monitoring and recording of the sleep sessions. The camera was mounted at a height of 2 m on the wall at the foot of the subject's bed and the monitor was located near the EEG machine.

A custom-built millisecond clock/sound card allowed precise presentation (and recording) of a 1,000 Hz tone to the subject. The tone was generated by a microprocessor-controlled oscillator on the clock card, which was inserted into an Apple IIe microcomputer. The Apple was programmed to generate tones lasting a maximum of 5,000 ms, with an intertone interval (ITI) ranging from 10 to 30 s. With a subject responding, the mean ITI was 21.2 s, SD = 5.6 s. Displayed on the VDU and filed for each trial were the following: response time (RT), ITI, total test time, tone number, and cumulative mean RT. The tone measured 5 dB(A) above background noise at the pillow, and was delivered through a 6-cm diameter speaker 1 m from the subject's head. Background noise averaged 27 dB(A). The term "response time" was chosen for the discrete behavioral response to tone to distinguish the task from reaction time measures. The latter term implies a level of readiness deliberately avoided in the instructions to subjects because the consequent increase in arousal would not be compatible with sleep.

A velcro strap held the RT miniswitch (sewn into a squash ball) in the palm of the sleeper's preferred hand so that it was comfortable to hold and constantly accessible. A weight of 125 g was required to activate the switch, which had a closure distance of 2 mm.

There were three microphones in the test chamber. Two were tiny (8 × 19 mm)

Electret condenser tie-clasp microphones; the one attached to the speaker housing where the tones were presented was connected to an audio channel on the VCR (to ensure distinct recording of tone onsets and offsets), while the other was taped on the midline of the subject's lower lip. The slightest vocalization triggered a voice-activated relay or voice key, the output of which was recorded on the EEG machine. A condenser microphone was suspended 0.2 m above the pillow, and was connected to the VCR, thereby enabling the experimenter to monitor and record the spoken (or whispered) SSS ratings with excellent fidelity.

A "deadman" (DM) switch was designed in an attempt to improve on the method used by Blake et al. (15)—dropping a piece of paper from between the subject's thumb and forefinger as an index of SO. The measure takes advantage of the fact that muscle tone diminishes during the S/W transition to a point where even slight pressure (on a switch) is difficult to maintain. Ability to attend to a task is presumed to decrease rapidly at this time. The DM device was constructed by attaching a miniswitch to the end of a rubber bicycle handle-bar grip. Following pilot studies, spring closure was found to require a movement of 3 mm and a constant force of 90 g to maintain contact. The DM switch enabled us to approximate a continuous behavioral reading of S/W on the EEG machine. To measure respiratory patterns, a dual graphite strain respiration transducer (16) was used.

Procedure: weekly schedule

Two new subjects were tested each week for 7 weeks. Prior to their test week, a 1-h pretesting session was held during which the volunteers familiarized themselves with the staff and the laboratory environment. Each participant was shown the recording equipment and then spent 10 min responding to the intermittent tones while lying in the test chamber. Every aspect of the subject's role was discussed before final scheduling took place. Volunteers had been asked to abstain from alcoholic beverages for 24 h and were not allowed caffeinated drinks during the study.

At about 2200 h, the two subjects arrived for the overnight sleep deprivation and subsequent SO testing. They had been advised to bring books, puzzles, knitting, and so forth to help pass the time from 2200 to 0730 h, during which no testing, other than half-hourly SSS ratings, was conducted. Completing the SSS throughout the deprivation period ensured that each participant was thoroughly familiar with the scale before SO testing began. The deprivation time was spent in a common room equipped with kitchen facilities, puzzles, a TV, and a number of VCR tapes of popular movies. A researcher was present at all times during the deprivation period to ensure that no napping occurred. The subjects ate breakfast at about 0700 h, then electrode application began at 0730. There were two sleep sessions per volunteer after the deprivation, one before and one after lunch, each lasting 60–90 min. The first SO session for the first subject began at about 1030 h.

Recording procedures

Sleep monitoring varied slightly from that described by Rechtschaffen and Kales (3). Two monopolar EEG electrodes were obtained from Cz and Pz—left mastoid electrodes—and horizontal and vertical EOGs were monitored from outer canthus and infraorbital placements, respectively, both referred to the mastoid. Midline (vertex) placements were better suited to detect the vertex sharp waves that are useful in defining Stage 1 sleep (17). Also, the Pz electrode was nearer to alpha-generating areas,

and the diminution of alpha is one of the traditional signs of SO. Bipolar submental EMG completed the usual sleep measures.

Experimental protocol

After electrode application, the volunteer was again taken to the EEG machine and shown that only a slight pressure was required to prevent the DM switch from indicating sleep. The respiration belts were attached and the electrode hook-up was completed. The squash ball/tone switch was secured to the preferred hand with velcro and the DM switch was fastened in a similar manner to the other hand. Also, a small microphone was taped to the subject's lower lip, on the midline to activate the voice key when SSS ratings were made. The functions of the tone button and DM switch were clarified, the 7 SSS scale points were read out, and the subject was asked for his or her present SSS scale rating.

The participant's duties were simple: (a) maintain pressure on the DM switch; (b) press the tone switch to terminate the tone whenever it was heard; (c) say a number from one to seven that represented the subject's current sleepiness level after each tone switch termination; and (d) relax and drift off toward sleep.

The volunteer was reminded that the experimenter would enter the sleep chamber on a number of occasions and call the person by name. He or she would be asked two questions at that point: "Were you awake or asleep as I entered the room?" "How many tones do you think you might have failed to respond to?" [The subject was not told that experimental arousals would be made following zero to seven consecutive failed responses to the tones. The number of failed responses for a particular arousal was determined randomly ($\bar{x} = 3.3$, SD = 1.7)]. After all last-minute questions had been answered, the recording session started.

Data preparation

It was decided that, wherever possible, the primary measurement of behavioral and physiological variables would be made at the onset of each tone throughout both test sessions for every subject. To accomplish that coordination, the following steps were taken. The EEG paper was scanned by an assistant who made a pencil mark above the Cz tracing to identify the onset of each tone. The records were then sleep scored using the Rechtschaffen and Kales (3) criteria, except that "point" (±2.5 s) S/W stage judgments were substituted for the usual practice of scoring more lengthy epochs. The paper was examined through a mask that concealed the three behavioral markers so that the experimenter was blind with respect to nonphysiological events. The longer of the a.m. or p.m. testing session was used for most calculations.

RESULTS

Chart recordings of SO

Before beginning detailed quantitative analyses of the following data, it may be helpful to undertake a visual inspection and description of two typical samples from the polygraphic records. The correlation matrices and ANOVA calculations presented later in this section confirmed the relationships first noticed by Ogilvie and Wilkinson (7). Two selections are displayed in Figs. 1 and 2 and described below.

Fig. 1 shows the progression from wakefulness to sleep. The photograph is of the final three tones in a subject's p.m. testing, a period that began with every index

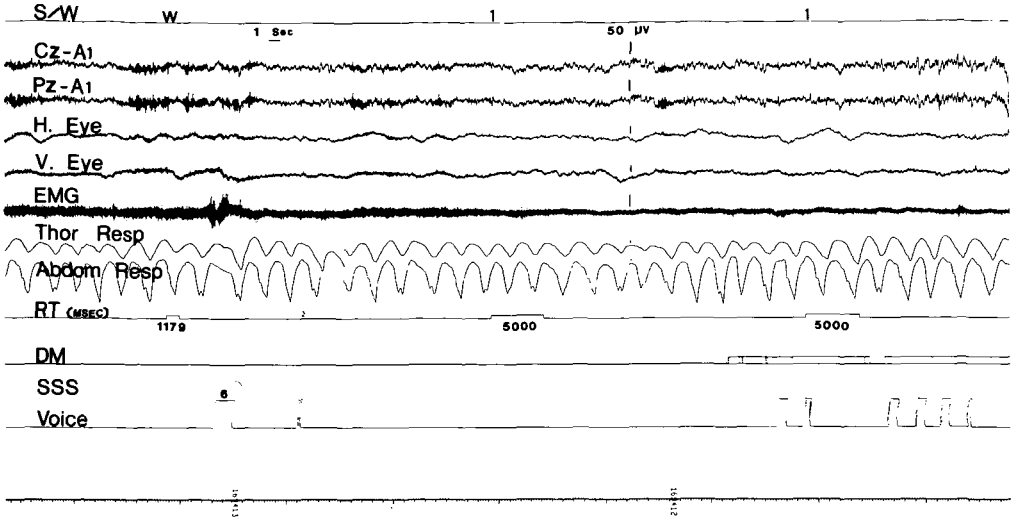

FIG. 1. Simultaneous physiological and behavioral measures trace the S/W transition. Point estimates of S/W stages are indicated above the CZ–A1 EEG tracings. Horizontal and vertical eye movements were recorded, as were submental EMG and thoracic and abdominal respiration. Response time is recorded in milliseconds, and deadman gripping or release (double line) can be seen. SSS verbalizations are indicated just above the output of the voice-activated relay that time-locked the SSS scores to the other measures.

pointing to wakefulness and ended in Stage 1–2 sleep. There was no button press or SSS rating given to the second tone; about 15 s after the tone ended the voice key was activated by the subject's snoring, which became very loud on the VCR (and evident on the voice key) by the end of Fig. 1. Generally atypical respiratory activity was associated with her snoring, which in turn may have been augmented by her obesity. It is also of interest to note that in this instance, the DM button was not released on the occasion of the first failed response, but did signal SO several seconds later.

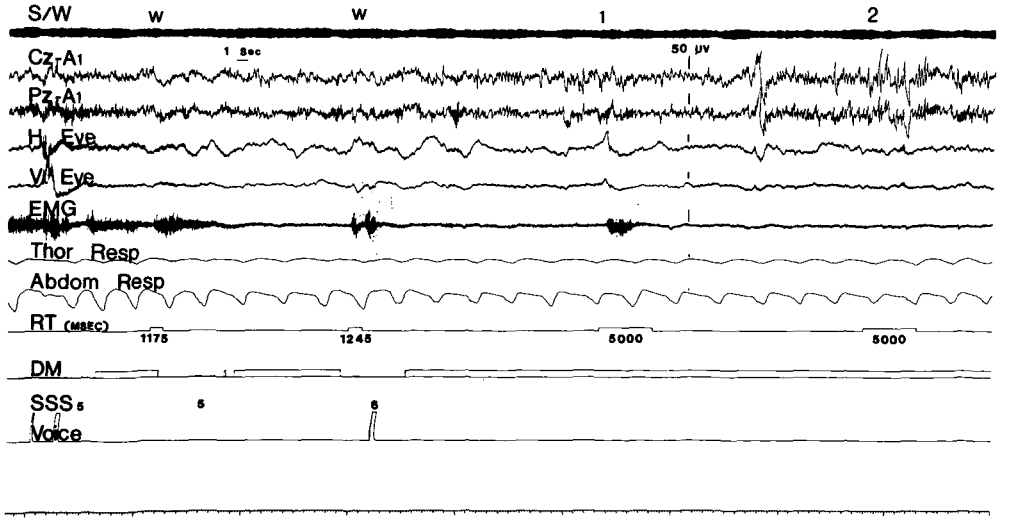

FIG. 2. Subjective and objective assessments of the W/S transition: no precise agreement. Legend as in Fig. 1.

Figure 2 again shows wakefulness to Stage 2 sleep, with waxing and waning alpha activity early on giving way to bursts of theta, and finally to K-complex and spindles as the sample ends. Changes in respiratory activity are also evident. The DM switch was released just after the last response to the tone, and subject 6 was aware of her increasing sleepiness as her altered SSS rating showed. Once again the behavioral measures combined to suggest that SO had occurred, but this time they identified the transition point as being in Stage 1, 10–25 s before the first K-complex located Stage 2 sleep. Figs. 1 and 2 show that the physiological and behavioral changes at SO are by no means identical in terms of there being a set sequence of changes that invariably leads to sleep.

Correlation matrices

At each tone onset the following measures were taken: RT, T, A, T/A (respiration) ratio, and SSS. "Point" EEG sleep staging was also performed. All possible pairs of correlations were then calculated, based on all tone-measurement points from each subject's longest recording session. Time was also included as a sixth variable to detect systematic changes in the other five variables during the course of the session. The passage of time was measured in intertone interval units of ~20 s ($\bar{x}$ = 21.2, SD = 5.6), which approximate an interval scale.

Table 1 presents a matrix of correlations among the above six measures. Although such a table could be added for the data of each subject, in the interest of brevity, Table 1 represents mean correlations over all 12 subjects together with standard deviations as some indication of the reliability of the coefficients. (Levels of significance cannot be reported for average correlations.) From Table 1, one can see that RT and respiratory measures produced a range of correlations; there was no relationship between T and RT, while moderate correlations of RT with A and T/A ratio were noted. Both SSS and time scores were also moderately but consistently related to RT. The three respiration measures intercorrelated among themselves reasonably well (r = −0.61 to r = 0.50). SSS related modestly to RT (r = 0.44).

TABLE 1. *Mean correlations among RT, respiration, SSS, and time: all trials to subject's first sleep onset*

	RT	T	A	T/A ratio	SSS	Time
RT	1	−0.01	−0.46	0.42	0.44	0.70
		(0.37)	(0.20)	(0.33)	(0.38)	(0.16)
T		1	0.43	0.50	−0.08	−0.10
			(0.39)	(0.38)	(0.36)	(0.40)
A			1	−0.61	−0.43	−0.59
			(0.39)	(0.34)	(0.33)	
Ratio				1	0.32	0.48
					(0.35)	(0.36)
SSS					1	0.72
						(0.53)
Time						1
Mean trial length	= 39.6					
SD	= 39.3					

r's transformed to Z-scores, averaged, and reconverted to r.
Standard deviations appear in brackets.
RT, response time; T, thoracic; A, abdominal; T/A ratio, respiration ratio; SSS, Stanford Sleepiness Scale.

Another phenomenon noticed during testing was that volunteers tended to change their SSS ratings much more frequently early in a session and were inclined to give a rather constant sleepiness assessment later on. On inspection of the data, other measures seemed to show the same tendency, so the above correlational analyses were performed on all variables only for the period up to the end of the first SOP. Inclusion of data obtained subsequent to initial SO lowered the correlations.

RT, respiration, and SSS as a function of EEG S/W stages

Having inspected the S/W traces in Figs. 1 and 2, one might wonder whether any or all of the measures shown could discriminate between EEG-based S/W stages. To answer that important question, data from each volunteer's longest test session (either a.m. or p.m.) were sorted according to EEG S/W state. Means were calculated for each dependent variable per subject per state and were in turn entered into the five repeated measures analyses of variance (ANOVAs) summarized below. EEG S/W state was restricted to Awake, Stage 1, and Stage 2 sleep, as there were insufficient instances of Stages 3, 4, and REM sleep to allow stable means to be computed. The mean numbers of observations contributing to the values entered for each subject were as follows: Awake, 83.6 (SD = 37.8); Stage 1, 47.7 (SD = 12.9); and Stage 2 sleep, 26.7 (SD = 13.1).

Table 2 displays the means and treatment F-ratios for the RT, respiration, and SSS measures in EEG Stages W, 1, and 2. With the exception of thoracic respiration, all the dependent measures showed significant differences among the EEG arousal states. The RTs changed as one would expect ($F = 144.4$, d.f. = 2,22, $p < 0.01$), with mean responses fastest during waking, slower in Stage 1, and slowest during Stage 2 sleep. Newman-Keuls a posteriori testing confirmed that mean RT was significantly different in each of the three states. Because the distribution of RT scores was skewed, a nonparametric ANOVA was also conducted. Over all three S/W stages, the Friedman Statistic (FS) was highly significant: $FS = 24.0$, d.f. = 2, $p < 0.001$. Since all subjects in this study also responded most rapidly in wakefulness and most slowly during Stage 2 sleep, the two two-groups comparisons between wakefulness and Stage 1, and between Stages 1 and 2, yielded identical statistics ($FS = 12.0$, d.f. = 1, $p < 0.001$). These findings confirm the parametric analyses reported earlier.

TABLE 2. *Repeated measures ANOVAS: RT, respiration, and SSS as a function of EEG S/W stage*

	Treatment means			
Dependent variable	Awake	Stage 1	Stage 2	F
RT	1371	3073	4804	144.43[a]
T respiration	6.72	6.55	8.18	1.14 NS
A respiration	7.59	5.20	4.56	16.63[a]
T/A ratio	0.78	1.03	1.20	5.58[b]
SSS	5.84	6.24	6.30	20.24[a]

[a] $p < 0.001$; [b] $p < 0.05$.
d.f. = 2,22 for all above F-tests.
Twelve subjects; mean score per arousal condition entered for each.
Underlined means not significantly different by Newman-Keuls analysis.
Abbreviations as in Table 1. S/W, sleep/wakefulness.

Three respiration measures—A, T, and T/A, as defined by Naifeh and Kamiya (10)*—will be examined next. There were no differences across EEG S/W stages among overall treatment means on the thoracic amplitude measure ($F = 1.13$, d.f. = 2,22, NS), and even the ordering of group means did not progress incrementally from wakefulness to sleep. Abdominal amplitude was the most sensitive respiratory index ($F = 16.63$, d.f. = 2,22, $p < 0.001$). Newman-Keuls analysis revealed that the significant change was a decrease in amplitude from wakefulness to Stage 1 sleep. Only a small, insignificant reduction in amplitude occurred between EEG Stages 1 and 2. Although the T/A ratio is useful because it is independent of respiratory amplitude, it did not provide as strong a measure of EEG S/W change as did relative changes in abdominal amplitude.

Finally, in the SSS ANOVA, yet another set of means differed significantly ($F = 20.24$, d.f. = 2,22, $p < 0.001$). Here too, the Newman-Keuls a posteriori testing confirmed that waking scores were lower than those accompanying both Stage 1 and Stage 2, and the latter two stages did not differ in mean SSS rating.

The DM switch—a continuous behavioral monitor

One of the advantages of using a measure like the DM switch is that it permits continuous monitoring of behavior to take place. The results of this monitoring were analyzed in several ways. First, comparisons were made between the DM and RT behavioral measures. Whether the switch was open or closed (depressed) was recorded for every RT (dichotomized as <5,000 or = 5,000 ms), and 2 × 2 contingency tables were constructed for both a.m. and p.m. sessions. Chi-square analyses were computed and phi and kappa coefficients obtained in an effort to achieve some comparability between the binary information provided by the DM and the other dependent variables. From Table 3, it can be seen that significant associations between the two behavioral measures were obtained for most subjects, thereby demonstrating the usefulness of the DM measure. The phi coefficients averaged 0.32, showing a moderate relationship between the two behavioral variables. In general, the chi-square and kappa coefficients were in agreement in identifying nonrandom distribution of tallies in the 2 × 2 contingency table. But on two occasions where chi-squares showed an overall pattern of significance, the kappa index of agreement between the DM and RT measures was not significant. The kappa statistic is the more appropriate of the two, being both more conservative and more precisely aimed at assessing the agreement diagonal of the 2 × 2 table, which reflects behavior in the expected direction, namely, that when the DM is on, behavioral responses will occur, and when it is released, they will not occur.

The next DM analyses were based on 2 × 3 chi-square tables with three EEG-defined levels of arousal—Stages W, 1, and 2—orthogonal to the binary DM index. Significant chi-square values were found for a majority of the volunteers, thus demonstrating a significant association between DM release and S/W stage.

Testing two definitions of sleep onset

An analysis with interesting theoretical implications was conducted (Table 4) in which the two most common EEG definitions of sleep were compared against the DM

* Respiration ratio (T/A) is an abbreviation of the formula $(T - A)/[(T + A)/2] + 1$, which appears in Naifeh and Kamiya and is used as a measure of respiratory activity that is independent of respiratory volume.

TABLE 3. *Phi coefficients and chi-square analyses of behavioral responses: 2 × 2 contingency tables—RT and "deadman"*

Subject	Phi	a.m. chi-square	Kappa	Phi	p.m. chi-square	Kappa
1	0.14	2.84 NS	0.13 NS	0.12	1.95 NS	0.10 NS
2	0.04	0.20 NS	0.02 NS			
3	0.41	12.19[a]	0.40[c]	0.34	18.34[a]	0.28[b]
4	0.18	3.66 NS	0.13 NS	0.37	11.47[a]	0.33[b]
5	0.60	52.50[a]	0.58[a]	0.29	14.17[a]	0.21 NS
6	0.04	0.22 NS	0.03 NS	0.42	14.98[a]	0.30[c]
7				0.27	14.43[a]	0.16 NS
8	0.33	18.64[a]	0.33[b]	0.12	1.38 NS	−0.07 NS
9	0.60	37.40[a]	0.60[a]	0.63	34.79[a]	0.61[a]
10	0.49	38.53[a]	0.38[a]	0.30	14.85[a]	0.30[a]
Total		166.18[a]			126.27[a]	
Mean	0.313			0.31		
SD	0.26			0.16		

[a] $p < 0.001$; [b] $p < 0.01$; [c] $p < 0.05$.
d.f. = 1 for each individual chi-square test.
d.f. = 9 for summation of individual chi-square tests.
RT: <5,000/5,000 ms.
Deadman: on/off.

data. To contrast these two definitions, the EEG data were dichotomized at two different points. For the standard definition of sleep (3), the dichotomy was between EEG Stage W and EEG Stages 1 + 2 sleep. For the definition of sleep according to Johnson (6), the dichotomy was between wake, which included Stages W and 1, and sleep, which was reflected by Stage 2 EEG. Chi-square scores were larger (Wilcoxon, $p < 0.05$, one-tailed) for the first than for the second definition of SO, respectively. The phi and kappa indices did not distinguish as well between the two definitions of sleep, although there were more individual kappa coefficients in favor of the first definition

TABLE 4. *Kappa and phi coefficients and 2 × 2 chi-square analyses: Sleep/wake versus "deadman" response*

Subject	Phi	Chi square (Awake = W; sleep = 1 + 2)	Kappa	Phi	Chi square (Awake = W + 1; sleep = 2)	Kappa
1	0.28	12.15[a]	0.23[b]	0.24	9.18[b]	0.24[c]
2	0.03	0.17 NS	−0.01 NS	0.08	1.01 NS	0.06 NS
3	0.31	16.26[a]	0.19[b]	0.36	20.35[a]	0.35[c]
4	0.34	9.65[b]	0.27[b]	0.24	4.67[c]	0.23 NS
5	0.26	11.37[a]	0.20[c]	0.28	14.15[a]	0.24 NS
6	0.45	16.76[a]	0.34[c]	0.23	4.33[c]	0.10 NS
7	0.59	67.27[a]	0.51[a]	0.26	13.77[a]	0.13 NS
8	0.39	26.40[a]	0.31[a]	0.26	11.98[a]	0.26[c]
9	0.51	27.36[a]	0.51[a]	0.49	25.19[a]	0.45[a]
10	0.48	36.88[a]	0.47[a]	0.23	8.70[b]	0.18[c]
Total		224.27[a]			113.33[a]	
Mean	0.36			0.27		
SD	0.16			0.11		

[a] $p < 0.001$; [b] $p < 0.01$; [c] $p < 0.05$.
d.f. = 1 for each individual chi-square test.
d.f. = 10 for summation of individual chi-square tests.

than the second. From the aforementioned data it would seem that the DM response tends to differentiate S/W better when Stage 1 is grouped with Stage 2 as "sleep" than when Stage 1 is added to Stage W.

Table 5 shows the results of a similar comparison, but one in which RT replaced DM. Trials where no response was made to the tone were separated from those where any RT of <5,000 ms occurred using a computer sort. Significant chi-squares and kappas were found for all subjects by both SO definitions, and substantial phi coefficients were obtained as well.

Intersubject variation in responsiveness is the focus of data summarized in Table 6. Failure to respond to the tones as a function of sleep state was examined and expressed as a percentage for each subject, and the ratios of failed responses to number of tones presented per S/W state were also noted. Examination of the latter shows each subject's responsiveness quite clearly. It will be recalled that a similar pattern of response failure was reported by Ogilvie and Wilkinson (7), although the proportion of response failure was slightly higher in the present investigation. Brief reference to individual variability in this table may be relevant to later discussions of the issue of sleep threshold. Two of the most extreme and opposite response patterns will be considered. Subject 8 responded to all 96 tones presented while he was in EEG-defined wakefulness; he missed only five of 53 tones in Stage 1, and failed to respond to all 23 tones while in Stage 2. For him, behaviorally defined SO would begin at or just before the beginning of Stage 2 sleep. On the other hand, subject 9 failed on 4 of 39 tones delivered while awake by EEG standards, missed 27 of 32 received in Stage 1, and did not respond to another 33 of 34 tones while in Stage 2 sleep. Her threshold could be said to occur much earlier, soon after she entered Stage 1 sleep. Considered a little differently, if threshold is taken as that point at which responding occurs 50% of the time, then none of these sleep-deprived individuals crossed that line while in wakefulness, but every

TABLE 5. *Kappa, phi coefficients and 2 × 2 chi-square analyses: Sleep/wake versus RT*

Subject	Phi	Chi square (Awake = W; sleep = 1 + 2)	Kappa	Phi	Chi square (Awake = W + 1; sleep = 2)	Kappa
1	0.57	50.27[a]	0.50[a]	0.63	60.65[a]	0.61[a]
2	0.47	35.28[a]	0.36[a]	0.62	60.64[a]	0.60[a]
3	0.72	81.52[a]	0.68[a]	0.49	37.57[a]	0.38[a]
4	0.63	47.30[a]	0.61[a]	0.64	48.11[a]	0.57[a]
5	0.60	62.21[a]	0.56[a]	0.62	65.84[a]	0.59[a]
6	0.74	79.41[a]	0.71[a]	0.65	60.75[a]	0.61[a]
7	0.42	33.59[a]	0.34[b]	0.58	65.24[a]	0.57[b]
8	0.50	42.25[a]	0.39[a]	0.89	136.54[a]	0.88[a]
9	0.80	66.99[a]	0.80[a]	0.51	27.54[a]	0.43[a]
10	0.51	42.88[a]	0.45[a]	0.48	36.58[a]	0.36[b]
11	0.50	48.26[a]	0.41[a]	0.83	130.35[a]	0.81[a]
12	0.62	64.17[a]	0.56[a]	0.62	62.84[a]	0.58[a]
Total		654.13[a]			792.65[a]	
Mean	0.59			0.63		
SD	0.12			0.12		

[a] $p < 0.001$; [b] $p < 0.05$.
d.f. = 1 for each individual chi-square test.
d.f. = 12 for summation of individual chi-square tests.

TABLE 6. *Individual differences in failed responses per sleep/wake stage*

Subject	Awake		Stage 1		Stage 2+	
	%	FR/total	%	FR/total	%	FR/total
1	1.5	1/67	34.6	18/52	85.7	30/35
2	0.0	0/60	24.2	16/66	79.4	27/34
3	0.0	0/75	60.0	39/65	100.0	19/19
4	11.4	4/35	56.4	22/39	100.0	44/44
5	4.4	4/91	39.1	18/46	86.1	31/36
6	1.6	1/61	57.1	24/42	95.2	40/42
7	1.2	2/164	13.0	3/23	71.4	5/7
8	0.0	0/96	9.4	5/53	100.0	23/23
9	10.3	4/39	84.4	27/32	97.1	33/34
10	2.0	2/102	34.0	18/53	100.0	7/7
11	0.0	0/137	12.2	5/41	100.0	12/12
12	0.0	0/75	40.7	24/59	90.6	29/32
Mean	2.7		38.8		92.1	
SD	4.0		22.7		9.6	

FR, failure to respond.

one had done so by the time Stage 2 sleep had developed. By this criterion, SO can be said to occur at some point during Stage 1 sleep.

Subjective estimates of sleep and wakefulness

In addition to the SSS scores, which were a major measure in this study, other assessments of subjective data were conducted. An index was constructed from responses to inquiries about whether volunteers had been awake or asleep just before the experimenter aroused them. Table 7 shows those responses for each subject and also a Sleepiness Ratio, this being the number of "asleep" reports divided by total arousals. The mean SSS value is also shown for each subject. The correlation between the Sleepiness Ratio and SSS ratings was as follows: $r = 0.53$, d.f. $= 11$, $p < 0.1$, which is suggestive of a relationship, but is not quite significant. Two other correlations are

TABLE 7. *Individual differences in subjective estimates of sleep and wakefulness*

Subjects	No. experimenter arousals	No. "awake"	No. "sleep"	Sleepiness N ratio	SSS
1	34	8	26	0.76	7.08[a]
2	21	8	13	0.62	5.89
3	21	8	13	0.62	5.30
4	26	1	25	0.96	6.53
5	33	7	26	0.79	5.82
6	36	9	27	0.75	6.12
7	3	3	0	0.00	4.16
8	14	11	3	0.21	6.70
9	24	0	24	1.00	6.37
10	17	8	9	0.53	6.25
11	6	5	1	0.17	5.94
12	25	6	19	0.76	6.30
Mean	21.7	6.2	15.5	0.60	6.04[b]
SD	10.4	3.3	10.4	0.32	0.75

The sleepiness ratio is defined as number of "sleep"/total arousals.
[a] This patient used "8" as a new level of sleepiness on the SSS.
[b] Point 6 on the SSS is defined as "sleepiness; prefer to be lying down; fighting sleep; woozy."

worth noting. The number of experimental arousals was not related to the number of instances when the volunteer perceived himself to be awake ($r = 0.17$, d.f. = 11, NS), but there was a strong relationship between the number of arousals by the experimenter and reports of having been asleep ($r = 0.95$, d.f. = 11, $p < 0.001$).

It will be recalled that subjects were aroused after zero to seven consecutive response failures, so it was of interest to determine whether reports of sleep or wakefulness were associated with different numbers of failed responses. As expected, longer trains of response failure were associated with reports of sleep rather than of wakefulness ($t = 2.54$, d.f. = 9, $p < 0.05$; awake $\bar{x} = 2.6$, SD = 0.97; asleep $\bar{x} = 4.06$, SD = 1.53). This finding shows that there is some overall accuracy in subjective awareness of S/W, but individual variance in this ability is very high.

A number of novel pieces of information regarding the objective identification and subjective perception of SO have been presented. Their theoretical and practical significance will be examined in the following pages.

DISCUSSION

Simultaneous recording of RT, EEG, respiration, DM, SSS, and subjective S/W estimates has highlighted the complexities of the process of falling asleep and has provided a data base from which to begin to integrate these findings in a new model of the SO process.

Sleep/wake stage differences: Cues to SO?

The RT ANOVA was again unequivocal in discriminating among the three EEG stages analyzed. Responding was slower (and less frequent) in Stage 1 than in wakefulness, and was significantly slower again in Stage 2 than when measured during Stage 1. This would indicate that there are discernable differences in responsiveness among the three stages, which in turn suggests that different levels of arousal are identified by standard categories of Wakefulness, Stage 1, and Stage 2 sleep. This important finding replicates an earlier study by Ogilvie and Wilkinson (7).

The discovery that A amplitude decreased significantly from wakefulness to Stage 1 sleep suggests that the changeover in respiratory control systems that is said to occur at SO (19) takes place during the transition between those two stages. The T/A ratio data confirmed the A analyses. Taken together, it seems that the important respiratory amplitude changes are brought about by the decreases in abdominal activity at SO. This agrees with the findings of Timmons et al. (8) and Naifeh and Kamiya (10).

In the progression from wakefulness through Stage 1 to Stage 2 sleep, only the change from Wake to Stage 1 was accompanied by a significant increase in SSS-measured sleepiness. These data show that near-maximum levels of sleepiness are attained by the time Stage 1 sleep is underway. This is a particularly interesting finding because often the perception of sleep does not occur until Stage 2 has been attained (19). However, since nonresponses occurred frequently during Stage 1 and predominated during Stage 2, SSS data were severely attenuated at those levels of arousal.

Together, the four analyses that showed EEG sleep stage differences shed more light on the issue of when SO can be said to take place. Studies on arousal thresholds have suggested, and these data confirm, that there is a lowering of arousal from wakefulness to Stage 2 (20,21), but few have examined the important events that take place before Stage 2 is fully developed. In all four of the ANOVAs in which significant differences

were obtained, a posteriori comparisons showed that significant changes had occurred between mean activity in wakefulness and in Stage 1 sleep. Only the RT index was sensitive to further decreases in arousal occurring between EEG Stage 1 and 2.

From the aforementioned comparisons, it would seem that no simple picture can be formed in the shape of the SO curve. It seems that different mechanisms shift at different times and rates as arousal levels begin to change, suggesting that the onset of sleep might best be viewed as occurring over a more protracted period than is usually considered.

Intercorrelations among measures

Systematic, frequent, and simultaneous measurement of five variables, for which ordinal scaling can be assumed, provides an ideal opportunity to examine the intercorrelations among them. Consistent moderate relationships between RT and respiratory and subjective estimates of sleepiness (SSS) as SO approaches are reported here for the first time. Earlier (7), EEG sleep stages and respiration were found to be related to RT. Here it was noted that SSS scores were associated with RT and respiratory measures consistently and significantly for most subjects, particularly if only data to initial SO were considered. After initial sleep, the SSS was unable to identify further levels of sleepiness, if these indeed exist. However, we do have evidence for the first time that a validated subjective index (the SSS) can correlate with other behavioral and physiological variables and often be as useful as they are in identifying movement along the arousal continuum towards early phases of sleep.

DM and RT responses at SO

In a classic study, Max (22) measured reductions in muscle tension at SO. On the strength of that finding, Blake et al. (15) used the dropping of a light object as a sign of sleep, and the DM switch was constructed to take advantage of both discoveries. The contingency table analyses of RT and DM data showed significant agreement for most subjects (eight out of 10). For seven of those same eight volunteers, analyses of DM data showed that differential responding occurred in the three S/W categories. Thus, the DM measure was useful since discriminations could be made with it between S/W according to both behavioral (RT) and EEG-based definitions of sleep. These findings agree with those of Perry and Goldwater (23), who used a passive behavioral system similar to the DM switch to assess SO. However, they anchored their instrument to changes in alpha level rather than to both active behavioral (RT) and EEG changes, as was achieved in the present report.

When the standard (3) definition of SO occurring at the beginning of Stage 1 was contrasted with the more conservative Stage 2 SO of Johnson (6), DM activity differentiated more clearly according to the Stage 1 definition of onset. Therefore, it would appear that the threshold of muscular relaxation required by the DM task occurred relatively early on in the descent down the arousal dimension.

Similar contingency analyses on RT data were very clear; RT scores differentiated S/W very well for both EEG definitions and on all 12 subjects. Thus, on a more active behavioral task, in which responses were required to specific stimuli, different conclusions with regard to levels of arousal could be drawn. In fact, in reference to Table 6, it will be recalled that the reorganization of RT as failed response percentages per sleep stage presents a strong case for viewing Stage 1 as the transition stage; undeniable signs of wakefulness and sleep can be found for every subject in Stages Wake and 2, respec-

tively, whereas for Stage 1, failed response percentages varied from nine to 84 and averaged 39%.

The SSS and sleep

The paradigm adopted here facilitated closer comparison among SSS and behavioral and physiological variables as SO approached than has been achieved in any other study. It was shown earlier that subjective sleepiness increased significantly as subjects moved from Stage Wake to Stage 1, and that no further significant increases in sleepiness occurred when Stage 2 was entered. The Stage Wake–1 changes provide important verification that perceived sleepiness increases as SO approaches. Significant correlations between SSS and RT and A respiration underscore that point. While these relationships are instructive, adapting the SSS to expand its ability to differentiate among extreme levels of sleepiness that accompany or follow SO would be very useful for both researchers and clinicians.

Subjective measures of sleep and SO

There was a high correlation between the number of behaviorally defined SOs that were terminated by the experimenter and the number of subjective reports of having been asleep, suggesting that the perception of sleep tends to increase as a function of the number of entries into sleep. Since total recording time was similar for all subjects, this would indicate that the shorter latency SOs are more likely to be perceived as sleep. These findings are consistent with reports (22,24) in which it has been shown that the perception of SO often lags several minutes behind objective SO indices.

The idea of identifying an SOP as distinct from both wakefulness and sleep was set forth by Ogilvie (24) and Ogilvie and Wilkinson (7). Data from the present investigation certainly support this concept. The SOPs can be identified quite precisely by behavioral criteria and satisfactorily by electrophysiological and subjective criteria. Using RT to define an SOP, increases in mean RT latency or decreases in the consistency of response rate or frequency locate the beginning of the interval, while RT cessation denotes the end of the SOP and the beginning of sleep. A roughly similar period was described as Stage 1 sleep by the Rechtschaffen and Kales (3) standard, but present data suggest that the SOP extends beyond Stage 1 in both directions. The full range of the SOP may best be identified if measurement begins during stabilized wakefulness and continues until slow wave sleep has been established. Whenever subjective SO is measured, the SOP might usefully extend to include the subjective beginning of sleep, for RT failure and the first sleep spindle will often occur before SO is perceived (22, present study).

Implications of an SOP for understanding sleep mechanisms

We previously argued that falling asleep does not take place at a precise point in time, but is better conceived of as occurring during an identifiable time period—the SOP. There are important implications of such a view for the understanding of the CNS systems whose activities are demonstrably different in waking and in sleep. There is indirect evidence in the present study that is consistent with current thinking that virtually all major brain regions undergo significant changes in activity as mammals shift from wakefulness to sleep. The abrupt changes in relative respiratory amplitudes early in the SOP are indicative of fundamental changes in brainstem control at SO. The increasing frequency with which people fail to detect simple tones during this period is

consistent with diminished cognitive functioning and decreased sensitivity to external stimulation, suggesting cortical (frontal?) and brainstem (plus possible midbrain and even spinal) involvement. The lapses in attention or fluctuations in arousal level that are detected by the RT system also suggest the possibility that both sensory threshold elevations and lowered levels of arousal are occurring during the SOP; here too lower centers are implicated. The latter is consistent with the growing importance of brainstem activation known to take place as sleep is entered. The passive release of the DM switch reflects the relaxation of both muscle tone and task awareness—the former is a precursor of sleep, while the latter helps define the state.

Finally, subjective changes in sleepiness as measured by the SSS showed that the most profound effects were visible early in the SOP, long before the perception of sleep took place. Our data show that there is an important gap between extreme sleepiness ("7" on the SSS) and the awareness of having been asleep, which is presently not adequately identifiable with present subjective assessments. This is not a criticism of the SSS, for that instrument was designed to measure sleepiness, not sleep perception. Nevertheless, improved scales of subjective experience vis à vis sleepiness and SO are required. In any event, it is clear that subjective awareness of sleep lags far behind behavioral and physiological evidence of the transition. For many purposes, such awareness might logically represent the end point of the SOP, for it is unlikely to occur before the appearance of virtually every other criterion for the identification of sleep. If changes in self-perceived arousal from drowsiness to sleep occur gradually throughout the entire SOP, perhaps they mirror such cortical changes as the slowing of the EEG from a predominance of beta, through the rise and fall of alpha, to and beyond the emergence of theta and the appearance of "sleep signs" such as spindles and K-complexes.

In conclusion, evidence has been presented that supports the adoption of the term "sleep onset period" to describe the transition from sleep to wakefulness. Behavioral, electrophysiological, and subjective evidence all point to the appropriateness of the SOP concept and to the refutation of the idea that Stage 1 represents sleep; Stage 1 events occur precisely in the middle of the SOP.

Acknowledgment: This study was conducted as part of a doctoral dissertation submitted to the University of Cambridge by R.D.O. Additional analyses were supported by an NSERC Grant awarded to R.D.O. We thank D. McDonagh for assistance with those analyses.

REFERENCES

1. Davis H, Davis PA, Loomis AL, Harvey EN, Hobart G. Changes in human brain potentials during the onset of sleep. *Science* 1937;86:448–50.
2. Dement WC, Kleitman N. The relation of eye movements during sleep to dream activity: an objective method for study of dreaming. *J Exp Psychol* 1957;53:339–46.
3. Rechtschaffen A, Kales A, eds. *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects.* Los Angeles: UCLA, Brain Information Service/Brain Research Institute, University of California at Los Angeles, 1968.
4. Agnew HW Jr, Webb WB. Measurement of sleep onset by EEG criteria. *Am J EEG Tech* 1972a;12:127–34.
5. Agnew HW Jr, Webb WB. Sleep stage scoring. *APA Journal Supplement Abstract Service.* Manuscript #293, 1972b.
6. Johnson LC. Are stages of sleep related to waking behavior? *Am Scientist* 1973;61:326–38.
7. Ogilvie RD, Wilkinson RT. The detection of sleep onset: behavioral and physiological convergence. *Psychophysiology* 1984;21:510–20.
8. Timmons B, Salamy J, Kamiya J, Girton D. Abdominal-thoracic respiratory movements and levels of arousal. *Psychon Sci* 1972;27:173–5.

9. Naifeh KH, Kamiya J. Changes of two respiratory parameters associated with sleep onset [Abstract]. *Sleep Res* 1978;7:71.
10. Naifeh KH, Kamiya J. The nature of respiratory changes associated with sleep onset. *Sleep* 1981;4:49–59.
11. Bonnet MH, Johnson LC. Relationship of arousal threshold to sleep stage distribution and subjective estimates of depth and quality of sleep. *Sleep* 1978;1:161–8.
12. Hoddes E, Zarcone VP, Smythe H, Phillips R, Dement WC. Quantification of sleepiness: a new approach. *Psychophysiology* 1973;10:431–6.
13. Kamiya J. Behavioral, subjective and physiological aspects of drowsiness and sleep. In: DW Fiske and SR Maddi, eds. *Functions of varied experience*. Homewood, Illinois: Dorsey Press, 1961.
14. Hersovitch J, Broughton R. Performance deficits following short-term partial sleep deprivation and subsequent recovery oversleeping. *Can J Psychol* 1981;35:309–22.
15. Blake H, Gerard RW, Kleitman N. Factors influencing brain potentials during sleep. *J Neurophysiol* 1939;2:48–60.
16. Ogilvie RD, Houghton D. A simple respiration transducer. *Electroencephalogr Clin Neurophysiol* 1982;54:347–8.
17. Fruhstorfer H, Paratanen J, Lumio J. Vertex sharp waves and heart action during the onset of sleep. *Electroencephalogr Clin Neurophysiol* 1971;31:614–7.
18. Phillipson EA, Murphy E, Kozar LF. Regulation of respiration in sleeping dogs. *J Appl Physiol* 1976;40:688–94.
19. Olmstead E, Hauri P, Percy L, Hellekson C. Subjective versus objective estimation of sleep onset in normal sleepers and in insomniacs [Abstract]. *Sleep Res* 1980;9:216.
20. Keefe FB, Johnson LC, Hunter DJ. EEG and autonomic response patterns during waking and sleep stages. *Psychophysiology* 1971;8:198–212.
21. Bonnet MH, Moore SE. The threshold of sleep: perception of sleep as a function of time asleep and auditory threshold. *Sleep* 15:267–76.
22. Max LW. An experimental study of the motor theory of consciousness. III Action-current responses in deaf-mutes during sleep, sensory stimulation and dreams. *J Comp Psychol* 1935;19:469–86.
23. Perry TJ, Goldwater BC. A passive behavioral measure of sleep onset in high-alpha and low-alpha subjects. *Psychophysiology* 1987;24:657–66.
24. Ogilvie RD. Behavioral and physiological investigations of sleep onset [Dissertation]. Cambridge, U.K.: University of Cambridge, 1985.