Available online at www.sciencedirect.com
ScienceDirect
Consciousness and Cognition 16 (2007) 241–254
Consciousness and Cognition
www.elsevier.com/locate/concog

The rotating spot method of timing subjective events

Susan Pockett [a,*], Arden Miller [b]

[a] Department of Physics, University of Auckland, Private Bag 92019, Auckland, New Zealand
[b] Department of Statistics, University of Auckland, Private Bag 92019, Auckland, New Zealand

Received 5 March 2006
Available online 17 Ocotber 2006

Abstract

The rotating spot method of timing subjective events involves the subject's watching a rotating spot on a computer and reporting the position of the spot at the instant when the subjective event of interest occurs. We conducted an experiment to investigate factors that may impact on the results produced by this method, using the subject's perception of when they made a simple finger movement as the subjective event to be timed. Seven aspects of the rotating spot method were investigated, using a factorial experiment. Four of these aspects altered the physical characteristics of the computer generated spot or clock face and the remaining three altered the instructions given to the participant. We found compelling evidence that one factor, whether the subject was instructed to report the instant when the finger movement was initiated or the instant when it was completed, resulted in a systematic shift in the response. Evidence that three other factors affect the observed variability in the response was also found. In addition, we observed that there are substantial systematic differences in the responses made by different subjects. We discuss the implications of our findings and make recommendations about the optimal way of conducting future experiments using the rotating spot method. Our overall conclusion is that our results strongly validate the rotating spot method of timing at least the studied variety of subjective event.
© 2006 Elsevier Inc. All rights reserved.

Keywords: Libet clock; Timing of subjective events

1. Introduction

The question of how to assign an objective time to internal subjective events has tested the ingenuity of scientists since the 19th century (Spence, Shore, & Klein, 2001). One well-known recent incarnation of Wundt's 'komplikationspendl' method of timing subjective events is the rotating spot method invented by Benjamin Libet and colleagues (Libet, Gleason, Wright, & Pearl, 1983). This method involves the subject's watching a spot of light (or in later versions sometimes a clock hand) rotating on an experimenter-generated clock face and then reporting the position of the spot at the instant when the subjective event in question occurred. Although Libet's version of the technique first appeared in the literature a little over 20 years ago, it presently remains essentially the only widely used method of timing subjective events.

* Corresponding author. Fax: +64 9 373 7445.
E-mail address: s.pockett@auckland.ac.nz (S. Pockett).

1053-8100/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.concog.2006.09.002

One problem with the rotating spot method is that the results it produces have been somewhat variable. Consider, for example, the various estimates it has produced of Libet's "time M"—the time at which a subject reports being aware of moving their finger to press a key or button. Libet's original averaged estimate of M (Libet et al., 1983) was −86 ms in relation to the onset of EMG activity. According to Haggard and Eimer (1999), EMG onset typically occurs 30–50 ms before the key or button is finally pressed, so in relation to the button press Libet's −86 ms would translate to about −126 ms. Later papers report M in relation to the button or key press at: −89 ms (Haggard & Eimer, 1999), -10 ± 34 ms (Tsakiris & Haggard, 2003), -1 ± 43 ms (Haggard & Clark, 2003), $+5 \pm 40$ ms (Wohlschlager, Haggard, Gesierich, & Prinz, 2003) and $+19.8 \pm 39$ ms (Sirigu et al. (2004)). The authors of these publications usually choose to render the 'accuracy' of the method relatively unimportant by using repeated measures/within subjects designs, which simply compare two experimental conditions within the same subject. However, it would be useful to know the source of the variation between the baseline results in different reports, so that the method might be refined to provide more reliable absolute timings for subjective events.

Thus, the aims of the present paper are to elucidate firstly what, if any, features of the rotating spot method might have played into the production of the fairly wide range of results in the literature and secondly how best to arrange the details of the rotating spot method to increase the accuracy and reduce the variability of subjective reports. To achieve these aims, we investigate the effects of altering seven different experimental factors on the results our subjects produced when asked to estimate Libet's time M.

2. Materials and methods

2.1. Participants

Twenty subjects, ranging in age from 19 to 67, each participated in one 20 min long experimental session. All but two of the subjects were members of the Physics Department of the University of Auckland. For the purposes of one of the analysis, subjects were divided into a "young" group of 10 subjects (9 male, age range 19–29) and an "old" group of 10 subjects (8 male, age range 46–67). The subjects were all essentially novices at the task—none received any extensive training prior to the experimental sessions reported here and most did no practice trials at all before to the first recorded trial. All subjects gave their informed consent as required by the University of Auckland Human Subjects Ethics Committee and all were rewarded at the end of their experimental session with one large chocolate fish.

2.2. Experimental procedure

The rotating spot clock was implemented in Delphi-6 on a laptop computer. The clock face consisted of a white circle against a black background, with 60 evenly spaced tick marks around the circle and numbers from 0 to 60 at intervals of five "clock-seconds". In the middle of the circle was a stationary white dot. The rotating spot (which could have various colors, sizes and rotation speeds, depending on the combination of factors currently being investigated) spun in a clockwise direction around the perimeter of the circle. The voluntary movement made by the subjects was to press the Enter key on the laptop. This had the effect of stopping the rotation of the spot, at a random time between 500 and 1500 ms after the key was pressed. The subject was then required to mouse-click on the circle at the position where they remembered the spot as being either at the instant they started to move their finger or the instant they finished hitting the Enter key (depending on the setting of factor 2 for that particular trial). The rationale for choosing what was essentially Libet's time M (rather than his time W) for study in this investigation was that there exists a convenient objective correlate of time M—the time at which the Enter key actually was pressed. The figures analysed in the present study were therefore the difference (in ms) between the objectively measured time of key press and the reported subjective time at which the subject thought they pressed the key (or in the case of factor 2 began to move).

2.3. Subject instructions

When subjects arrived, it was explained to them that they would be doing 4 blocks of 20 trials, with a general set of instructions that covered all of the blocks and additional sets of instructions that would be different for each block. The general instructions were that on each trial, the experimenter would start the spot rotating around the circle and the subject was to wait for one full rotation of the spot and then press Enter at a time of their own choosing, using whichever hand they preferred. It was emphasized that they were not to preplan when to press Enter—for example they should not decide that they would press the key when the spot reached a particular number on the circle. The phrase used was that the subject was in charge, not the clock. Subjects were told that after the Enter key was pressed, the spot would continue to rotate for a short but random period and then stop. The centre dot would then turn red, which was a signal that they should use the mouse to move the cursor to the position on the circle where the rotating spot was at the instant they pressed Enter. After these instructions had been given, the subject was asked to tell the experimenter what they understood that they would be doing, and any errors or omissions in understanding were corrected. When the general instructions were established, instructions specific to each of the four blocks of trials were given immediately before the block began (again with a request that the subject tell the experimenter what they were going to do). Specific instructions for the various factors investigated are given below.

2.4. Factors investigated

The effects of seven factors were investigated, with two settings for each factor (Table 1). The two settings for factors 1, 2 and 5 varied in the instructions given to the subject and for factors 1, 2, 3 and 7 varied in the physical parameters of the clock. The seven factors investigated and the rationale for studying each of them are as follows:

(1) Factor 1 is labelled Urge/Decision (Table 1). The two settings of this factor investigate possible differences between allowing an urge to move the finger to arise spontaneously (instructions: "Allow the key-press to happen of its own accord. Report the position of the spot when you first became aware of the urge or desire to move") and making a definite decision to act (instructions: "Make a definite decision to act now. Say to yourself NOW and then immediately press the key. Report the position of the spot at the instant you said NOW.") In this case, factor setting -1 corresponds to Libet's original instructions (Libet et al., 1983) while setting $+1$ is included to provide a baseline for possible future investigation of the idea (Mele, 2006) that there might be some meaningful subjective difference between the timing of an actual decision and a mere urge.

(2) The two settings of factor 2 (Start/End Movt) require the subject to report the position of the spot either at the start or the end of the movement—the instant at which they started to move their finger, or the instant at which they estimated the key press to be complete. This factor was included essentially as a test of the accuracy of the rotating spot method as a whole. If the method cannot distinguish between two events that must objectively be separated by at least a few tens of ms, it can not be considered very accurate in any objective sense.

Table 1
Factors investigated distinguishing features of the two levels for each factor are shown

Factor	Setting -1	Setting $+1$
X_1 Urge/Decision	Allow urge to appear spontaneously	Make definite decision to ACT NOW
X_2 Start/End Movt	Report start of finger movement	Report time keypress complete
X_3 Dark/Light Spot	Dark blue spot	Lime green spot
X_4 Little/Big Circle	Circle radius 1.3 cm	Circle radius 2.6 cm
X_5 Centre/Follow	Fix gaze on centre of circle	Follow spot round with eyes
X_6 Little/Big Spot	Spot diameter 2 mm	Spot diameter 3.5 mm
X_7 Slow/Fast Rotation	Rotation speed 3.19 cm/s (Period 2.56 s for small circle, 5.12 for large circle)	Rotation speed 6.38 cm/s (Period 1.28 s for small circle, 2.56 s for large circle)

(3) Factor 3 (Dark/Light Spot) investigates the effect of the visibility of the rotating spot. A dark blue spot was considered *a priori* to be less visible against a black background than a lime green spot and the subjective truth of this was commented on without prompting by a number of subjects. The rationale for including this factor was that it has been suggested (Klein, 2002; van de Grind, 2002) that the flash lag effect might have an influence on the results of the rotating spot method. It is known that the flash lag effect itself is modulated by the relative luminance of flash and moving spots (Lappe & Krekelberg, 1998; Purushothaman, Patel, Bedell, & Ogmen, 1998), so if the flash lag effect is involved here, there should be some difference between the results using a rotating spot with low luminance (or at least low visibility) and those using a rotating spot with greater luminance or visibility.
(4) Factor 4 (Little/Big Circle) investigates the possibility that the radius of the circle around which the spot rotates might affect the results of the method. Libet's original instructions required the subject to keep their gaze fixed on the centre of the clock face and note the position of the revolving spot using only peripheral vision. When this instruction is followed, it might be harder to report spot position accurately if the radius of the circle is larger.
(5) Factor 5 (Centre/Follow) requires the subjects either to keep their gaze fixed on the centre of the clock or to follow the spot round with their eyes. As mentioned above, Libet's original instructions were to keep the gaze fixed on the centre of the clock. However, there is a natural tendency for the eye to follow a moving object, and it is possible that some of the variation in the results of later investigations might be due either to a failure to emphasize the instruction to fix the gaze on the centre of the circle or a failure of subjects to follow this instruction.
(6) Factor 6 (Little/Big Spot) concerns the size of the rotating spot itself. There is a potential trade-off here between ease of visibility of the spot and the intrinsic temporal resolution of the method. With a rotation period of 2.56 s, each of the 60 marked "clock seconds" represents approximately 43 ms, so if the spot is so large that it spans a number of "clock seconds" the temporal resolution of the method for each trial would be significantly reduced. The spot diameters used in the present study were 2 and 3.5 mm. In this case the larger spot spanned two clock seconds.
(7) Factor 7 (Slow/Fast Rotation) investigates the effects of fast or slow rotation of the spot. Again the rationale was that there might be a trade-off between intrinsic temporal resolution and difficulty of pinpointing the position of the spot. For example, with a rotation period of 1.28 s, each "clock second" represents only about 21 ms, which implies potentially increased accuracy—but on the other hand, at such a fast rotation speed it may be harder to be accurate about the position of the spot at any given instant.

2.5. Experimental design

As our main goal was to study the seven treatment factors described in Table 1, we wanted an experiment which would give uncorrelated estimates of the effects of each of these factors. In addition, we wanted to be able to assess whether any of these factors interacted with the other factors—that is, whether the effect of any one factor depended on the setting of one (or more) of the other factors. The most commonly used type of design to achieve these aims is the factorial design. A full factorial design requires that every possible combination of factor levels is used and thus in our case would require that $2^7 = 128$ combinations or runs be tested. Instead, we decided to use a half-fraction of the factorial design, which reduced the number of runs to 64. The selection of fractions of factorial designs for two-level factors has been extensively studied and is covered in many texts on the statistical design of experiments—for example, see Wu and Hamada (2000, pp. 153–204). The half-fraction we used allowed uncorrelated estimates to be obtained for all seven main effects and all 21 two-factor interactions for the seven treatment factors.

The main effect of a two-level factor is defined as the difference in the population mean response between the two levels of that factor (averaged over all combinations of levels for the other factors). For example, the main effect for Little/Big Spot (X_6) is the mean response when a big spot is used minus the mean response when a little spot is used. For our experiment this main effect is estimated by the average of all responses where a big spot was used ($X_6 = +1$) minus the average of all responses where a little spot was used ($X_6 = -1$):

$$(\mathrm{Yave}|X_6 = +1) - (\mathrm{Yave}|X_6 = -1)$$

Two-factor interactions measure how much the effect of one factor depends on the level of another factor. For example, the Little/Big Spot by Slow/Fast Rotation (X_6:X_7) interaction measures how much the difference in the mean response between little spots and big spots changes when the fast rotation speed is used instead of the slow speed. This definition is mathematically equivalent to that obtained by interchanging the two factors. That is if Slow/Fast Rotation has a given impact on the effect of Little/Big Spot then Little/Big Spot must have exactly the same impact on the effect of Slow/Fast Rotation. For the current example the interaction is estimated by

$$\frac{[(\text{Yave}|X6=+1,X_7=+1)-(\text{Yave}|X6=-1,X_7=+1)-(\text{Yave}|X6=+1,X_7=-1)+(\text{Yave}|X6=-1,X_7=-1)]}{2}$$

The divisor of two is used so that the estimated two-factor interactions have the same standard deviation as the estimated main effects.

In our primary experiment, each of 16 subjects was used to test four of the treatment combinations during a single testing session. Table 2 gives the sets of treatment factor settings and the order in which these were tested in for the first two subjects. We were concerned that there could be systematic differences between subjects and between slots—i.e., that the response for a particular treatment combination may be affected by which subject was used and whether that run was tested first, second, third or fourth. Thus the experiment was "blocked" on both subjects and slots—i.e., runs were assigned to subjects and slots in a manner that minimised their impact on the estimated effects. For example, consider the estimated main effect of Little/Big Spot. For each subject, two of the runs used little spots and two used big spots. Thus any systematic effect due to a particular subject should have approximately the same impact on the mean for big spots as it has on the mean for little spots. As the estimated effect is the difference of these two means the effect of subject is canceled. Similarly, each slot was used for one quarter of the runs where a big spot was used and one quarter of the runs where a small spot was used. Thus systematic differences between slots should also be canceled. The design we used was able to make the estimates of all main effects and 16 of the 21 two-factor interactions insensitive to differences between subjects, and all of the estimates insensitive to differences between slots. An additional advantage of using subjects and slots as blocks is that we were able to test for both subject and slot effects.

In addition, we divided the subjects into two groups based on age. Our design allowed us to test for a main effect due to age but not for interactions involving age.

After analysing the results of the experiment described above, we decided to run a smaller follow-up experiment to clarify the results of the primary experiment. The follow-up experiment was a 16-run, full-factorial design for four of the original factors. We used four new subjects, each of whom tested four combinations of factor settings.

Table 2
Factor settings for Subjects 1 and 2 in primary experiment

Subject	Slot	X_1 Urge/ Decision	X_2 Start/ End Movt	X_3 Dark/ Light Spot	X_4 Little/ Big Circle	X_5 Centre/ Follow	X_6 Little/ Big Spot	X_7 Slow/ Fast Rotation
1	1	−1	1	1	−1	1	1	1
1	2	1	−1	−1	1	1	−1	1
1	3	1	−1	1	−1	−1	−1	−1
1	4	−1	1	−1	1	−1	1	−1
2	1	−1	1	−1	−1	−1	−1	1
2	2	1	−1	1	1	−1	1	1
2	3	1	−1	−1	−1	1	1	−1
2	4	−1	1	1	1	1	−1	−1

For each subject, 20 trials were conducted using the combination of factor settings in each of the four slots, making a total of 80 trials per subject. The −1 setting indicates that the first of the two options for that factor was used and the 1 setting indicates the second of the two options. Thus for Subject 1 Slot 1, the instructions to the subject were to allow the urge to move to arise spontaneously ($X_1 = -1$), report the time at which the movement ended ($X_2 = 1$) and follow the spot round with the eyes ($X_5 = 1$) and the clock consisted of a big ($X_6 = 1$), light colored ($X_3 = 1$) spot, rotating fast ($X_7 = 1$) around a small circle ($X_4 = -1$). Since the aim of this table is to illustrate the overall shape of the experimental design, the factor settings are left in the formal −1 or 1 format rather than being translated into actual instructions or clock parameters.

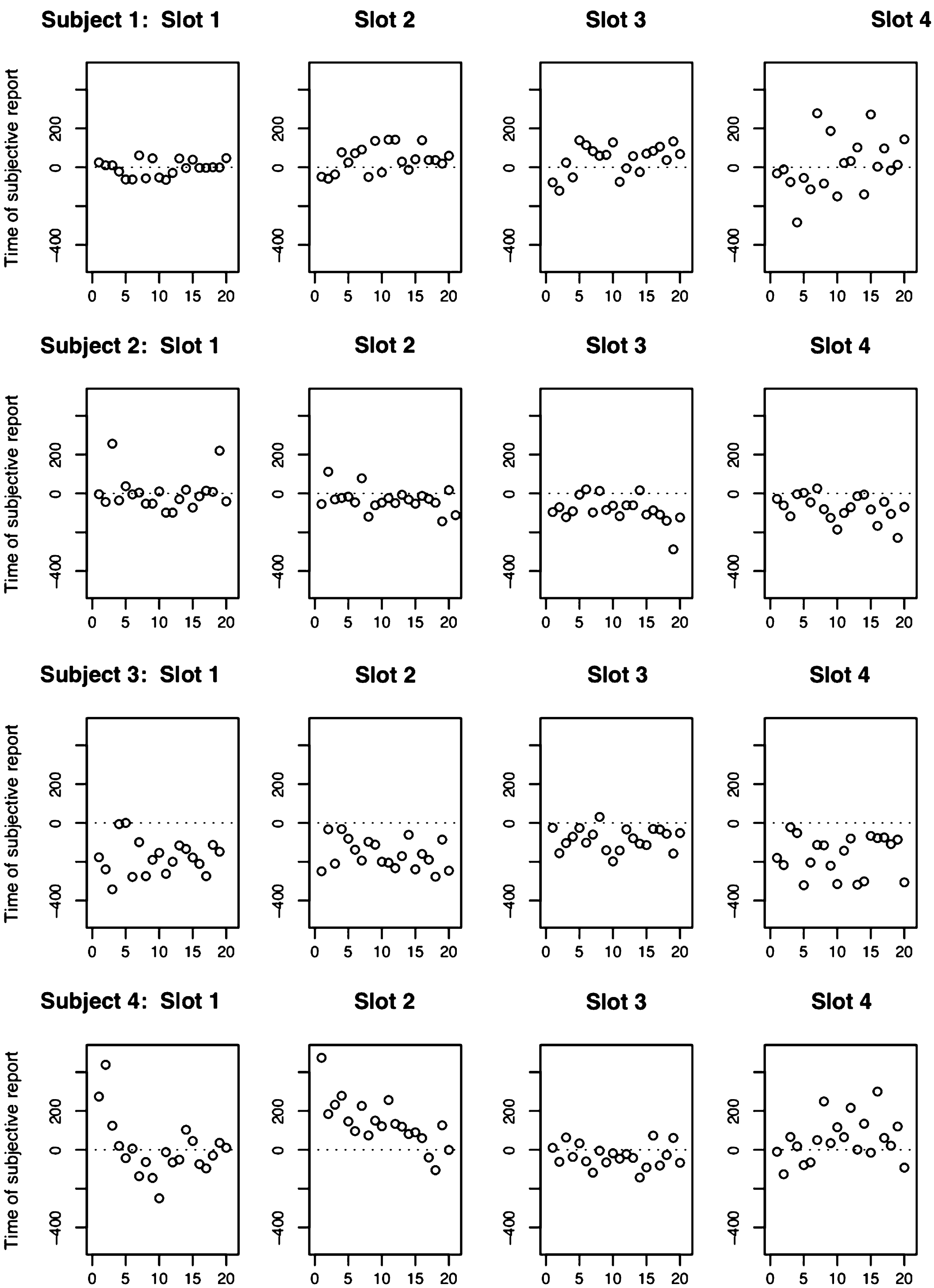

Fig. 1. Raw data for four subjects. Horizontal axis shows trial number. Vertical axis shows times (ms) of Libet's time M: zero is objectively measured time of key press.

2.6. Methods of analysis

For each slot of trials, the subject's report of spot position for that combination of instructions and clock parameters was measured approximately 20 times. Fig. 1 shows plots of the measurements for each of the four slots completed by four of the subjects in the primary experiment. Two analysands were chosen for each slot: one to describe the centre of the measurements and a second to describe the variability. Because subjects sometimes admitted at the end of a slot of trials to losing concentration on one or two trials and having essentially to guess where the spot was at the instant of interest, it was desirable to define these features in such a way that they would be insensitive to a small number of unusual observations. For example, in the plot for Subject 2 Slot 1 in Fig. 1, the two highest values appear to be distinct from the remaining points and may be due to momentary losses of concentration by the subject. Therefore, a trimmed mean was chosen as an appropriate estimate of the centre and the interquartile range (IQR) as the estimate of the variability of the measurements obtained for each block of trials. The trimmed mean was calculated by removing the largest 20% and the smallest 20% of the values and taking the mean of remaining values. The IQR is the difference between the 25th and 75th percentiles of the observations.

Both trimmed mean and IQR were analysed using ANOVA to test for main effects and two-factor interactions, based on the assumption that all interactions involving three or more factors are negligible. This is a standard method of analysing data from experiments that use fractions of a full-factorial design and is based on the empirical observation that two-factor interactions occur with much greater frequency than higher order interactions. For our design, the ANOVA table was split into three sections called strata (Subjects, Slots and Within), where each stratum had a different amount of variability associated with it. The Subjects stratum was affected by differences between subjects (but not slots), the Slots stratum was affected by differences between slots (but not subjects) and the Within stratum was not affected by either differences between subjects or differences between slots. As there was less variability in the Within stratum, tests made in this stratum were more sensitive than those made in either the Subjects stratum or the Slots stratum. For our design, the seven main effects and 16 of the two-factor interactions were tested in the Within stratum, while the remaining five two-factor interactions and the main effect for age are tested in the Subjects stratum.

To see whether there were significant differences between subjects and between slots, *F-tests* that compared the residual mean squares of the Subject stratum and of the Slot stratum with those of the Within stratum were used.

For each main effect and each two-factor interaction, the ANOVA table tests the hypothesis that the effect is really zero. The *p*-value can be interpreted as the probability of getting an estimate as far from zero as what was observed if the true effect was really zero. Typically, *p*-values of less than .05 are interpreted as providing reasonable evidence of a non-zero effect. However in our case, more caution was desirable since we were testing a large number of effects (seven main effects and 21 two-factor interactions), which means that if all the effects were really zero, it would not be surprising to find a few *p*-values $<.05$ just by chance. Thus, in addition to looking at the *p*-values for the individual tests, we also performed an adjustment for multiple tests using the multivariate *t*-method (Milliken & Johnson, 1992, p. 34).

Individual and simultaneous 95% confidence intervals were calculated for each of the main effects. For the individual intervals, it is possible to be 95% confident that any one of these intervals will contain the true value for that effect. For the simultaneous intervals, we can be 95% confident that all seven contain the true values. An individual 95% confidence interval is calculated as the estimate $\pm\, t_d\,(0.975) \times$ the standard error of the estimate, where $t_d\,(0.975)$ is the 97.5th percentile from at distribution with d degrees of freedom. To calculate the simultaneous confidence interval, the $t_d\,(0.975)$ is replaced by the 97.5th percentile from a 7-variate multivariate t distribution.

3. Results

3.1. Impact of the tested factors on central tendency

The ANOVA table for the trimmed means in the primary experiment indicated that three of the main effects and one two-factor interaction were significant at the 5% level. The significant effects were Start/End Movt ($p = .0025$), Dark/Light Spot ($p = .043$), Centre/Follow ($p = .016$) and the Start/End Movt by Little/Big Spot interaction ($p = .0083$). However, adjusting for multiple tests indicated that there was moderate evidence that Start/End Movt was significant (adjusted p-value = approx .05) but no evidence that any of the other three effects were significant (adjusted p-values $>.10$). When subjects were asked to report the start of their finger movement the average response was approximately 39 ms before the actual keypress whereas when asked to report the time the keypress was complete the average response was approximately 10 ms before the actual keypress.

To clarify the results from the initial trimmed mean analysis, a follow-up experiment was conducted to re-examine the effects of Start/End Movt, Dark/Light Spot, Centre/Follow and Little/Big Spot. In this experiment, the remaining factors were set to reduce variability (details to follow). Analysis of the trimmed means from the follow-up experiment confirmed that Start/End Movt was significant ($p = .034$) and that Dark/Light Spot, Centre/Follow and the Start/End Movt by Little/Big Spot interaction were not ($p = .51$, .73 and .93). In this second experiment, one subject was unusual in that her responses when asked to report the end of the movement were much higher than those of any of the other subjects tested. The analysis of the second experiment was therefore rerun omitting the results for this subject but gave the same conclusions. We conclude that our investigation does provide compelling evidence that factor Start/End Movt (the requirement to report either the start or the end of the movement) does affect the trimmed mean, but does not provide any such evidence for any of the other factors.

Table 3 shows the estimated main effect for each of the seven factors. The estimates for Start/End Movt, Dark/Light Spot and Centre/Follow are based on 19 subjects (all subjects except the unusual one from the second experiment) and the remaining effects were estimated using the 16 subjects from the first experiment. Table 3 also shows individual and simultaneous 95% confidence intervals for these main effects. The main message from Table 3 is that, while Start/End Movt was the only factor for which the main effect was statistically significant, several other factors could have effects that are almost as large. For example, the confidence interval for Centre/Follow extends from −2.3 to −36.3 as compared with the estimated effect for Start/End Movt of 32.4. This suggests that in a larger experiment, Centre/Follow might conceivably have demonstrated a significant effect. The direction of that effect is illustrated in Table 4, which shows the means of the observed trimmed means for the various combinations of Start/End Movt and Centre/Follow (for example, for reporting the start of the movement and focusing on the centre of the circle there would be 16 trimmed means—one for each subject—and the value in the table is the average of these). It can be seen that for both

Table 3
Estimates and confidence intervals for main effects

Factor	Estimated effect	95% Confidence interval individual	95% Confidence interval simultaneous
X_1 Urge/Decision	−2.1	(−19.2, 15.0)	(−26.2, 22.0)
X_2 Start/End Movt	32.4**	(15.4, 49.4)	(8.4, 56.4)
X_3 Dark/Light Spot	14.8*	(−2.2, 31.8)	(−9.2, 38.8)
X_4 Little/Big Circle	7.0	(−10.1, 24.1)	(−17.1, 31.1)
X_5 Centre/ Follow	−19.3*	(−36.3, −2.3)	(−43.3, 4.7)
X_6 Little/Big Spot	−6.1	(−23.2, 11.0)	(−30.2, 18.0)
X_7 Slow/Fast Rotation	11.6	(−5.5, 28.7)	(−12.5, 35.7)

Main effects and confidence intervals calculated as in text. Double asterisk indicates significance at the 0.001 level, single at the 0.05 level. Adjustments for multiple tests reduced the significance level of X_2 to 0.05 and suggested that X_3 and X_5 were not significant.

Table 4
Effects of factor 5 at different settings of factor 2

X_2 setting	X_5 setting	Mean
Report start of movement	Fix gaze on centre of circle	−36.4
Report start of movement	Follow spot round with eyes	−41.3
Report end of movement	Fix gaze on centre of circle	9.6
Report end of movement	Follow spot round with eyes	−29.5

Means shown are the simple means of each group of 16 individual trimmed means in the primary experiment.

settings of Start/End Movt, i.e., whether the subject was required to report the start or the end of the movement, the reported time is more negative when the subject follows the rotating spot around with their eyes than when they fix their gaze on the centre of the circle.

3.2. Impact of tested factors on variability within subjects

Fig. 1 shows raw data for each of the four slots or trials completed during the single 20-min session undertaken by four of the subjects. It can be seen that the variability of responses within individual slots was itself quite variable. It was not possible to do a statistical test on whether variability either increased or decreased in a regular way with time during each individual session, but no trends were obvious in the data.

On a statistical level, the ANOVA for the logged IQR's from the primary experiment showed that variability was affected by Little/Big Circle ($p = .0045$), Slow/Fast Rotation ($p = .0021$) and possibly Dark/Light Spot ($p = .0096$) and Urge/Decision ($p = .035$). Adjusting for multiple tests resulted in p-values of approximately .05 for Little/Big Circle and for Slow/Fast Rotation, approximately .10 for Dark/Light Spot and >.10 for Urge/Decision. For the follow-up experiment, the subject was instructed to make a definite decision to act now and a little circle and fast rotation speed were used—since the primary experiment indicated that these were the levels which should give lowest variability. Plots of the responses in the follow-up experiment showed that this strategy was successful in reducing variability.

3.3. Variability between subjects and overall central tendency

Fig. 1 illustrates the fact that some subjects (for example subject 3) returned mean response times quite different from those of other subjects. Statistically, we found extremely strong evidence of systematic bias, or differences between subjects with respect to their central tendency, independent of the effects of the factors being tested. The F-test of no differences between subjects from the trimmed mean ANOVA table resulted in a p-value of .00046.

To investigate further the differences between subjects, we divided the observed trimmed means into groups based on the level of Start/End Movt (since this was the only factor that had a significant impact on the trimmed mean). The experimental design dictated that for each subject, two observations were made for each level of Start/End Movt. These two observations were averaged over all subjects, resulting in a means ± SE of -36.8 ± 48.8 for reporting the start of the movement and of -9.2 ± 92.2 for reporting the end of the movement. Fig. 2 shows box plots for these two groups.

The box plots summarise the range of trimmed mean values observed across subjects. Each box plot has a dark line at the median observation, a box that extends from the 25th percentile to the 75th percentile, lines that extend to the most extreme points that are no more than 1.5 times the width of the box from the end of

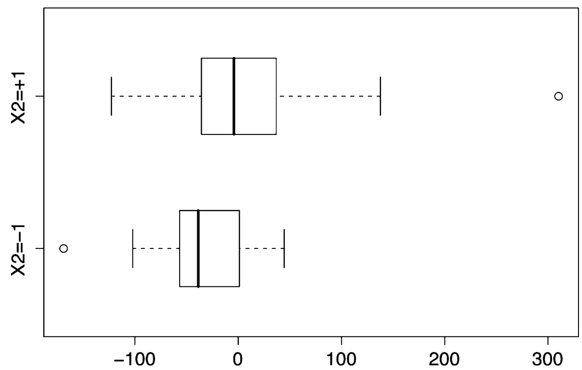

Fig. 2. Box plots of results for the two levels of factor 2 over all subjects. $X_2 = +1$ includes all trials where subject was instructed to report end of movement. $X_2 = -1$ includes all trials where subject was instructed to report start of movement. Heavy lines in middle of boxes are medians. Boxes extend from 25th percentile to 75th percentile. Dotted lines extend to most extreme points that are no more than 1.5 times the width of the box from the end of the box. Circles are outliers (observations outside these limits).

Table 5
Estimates of central tendency

Measure	Report start of movement	Report end of movement
Mean	−36.8 (10.9)	9.2 (20.6)
Mean (outliers removed)	−29.8 (8.9)	−6.2 (13.9)
Trimmed mean	−33.7 (10.4)	−5.2 (13.7)
Median	−38.6 (11.6)	−4.1 (11.2)

Estimates of central tendency are across subjects. Units are ms, with zero = actual time of key press (i.e., end of movement). Figures in brackets are standard errors of estimates. $n = 20$.

the box, and points for observations that are outside these limits (outliers). Thus for reporting the start of the movement ($X_2 = -1$), the middle 50% of subjects ranged from −54.3 to −0.1 with a median of −38.7.

Table 5 gives values for the mean, mean with outliers deleted, trimmed mean and median of the data. It can be seen that there is generally good agreement among these estimators, although the mean is very sensitive to the presence of outliers (note particularly the strong effect of one outlier in reporting the end of the end of the movement).

3.4. Effects of age

There was no significant effect of age on trimmed means ($p = 0.5$). For logged IQRs ($p = 0.06$) there was some evidence that the ‘‘old’’ group produced more variable results than the ‘‘young’’ group.

4. Discussion

4.1. Measures of central tendency

The first question that needed to be addressed in analysing our data was what measure of central tendency it was the most appropriate to use. The measure most commonly used in the literature on rotating spot experiments is the simple mean. However:

(a) the mean is really only a sensible measure for non-skewed distributions. For a skewed distribution, the median (the point which is below half the observations and above half the observations) is most sensible, because it is sensitive to the main part of the distribution and not the tails. In this respect the trimmed mean is a compromise between the mean and the median. Since our distribution is somewhat skewed, we considered either the median or the trimmed mean to be more appropriate measures than the mean.
(b) as well as being influenced by skewness, estimates of the mean are heavily influenced by outliers. Use of the either median or the trimmed mean almost eliminates the impact of outliers. In our data there were a number of outliers, which we inferred on the basis of subjective report by the subjects to be probably due to momentary losses of concentration. The presence of such outliers can be seen in the raw data for Subject 2, Slot 1 in Fig. 1 and their influence is reflected in the difference between the mean and the other measures for reporting the end of the movement in Table 5.

On these grounds, we chose—and it seems reasonable to recommend that future studies of this sort should also choose—to analyse either trimmed means or the medians of the data, not simple means.

4.2. Effect of reporting the start or the end of the movement

The only tested factor which had any significant effect on the trimmed means of our data was Start/End Movt. The two settings were instructions either to report the time at which the subject perceived that they started to move their finger to press the Enter key or to report the time at which they thought the pressed key hit the bottom of its range. Objectively, the time at which the movement started would be expected to be 30–50 ms before the time at which the key press was completed (Haggard & Eimer, 1999). On average, our subjects reported that they thought the end of the movement happened almost exactly when it objectively did occur (-4.1 ± 11.2 ms, median $\pm$ SE) and the start of the movement happened almost exactly 40 ms before the end of the movement as objectively measured (-38.6 ± 11.6 ms, median $\pm$ SE). This shows that the combination of the subjects' perceptions and the rotating spot method of reporting them proved in the case of our particular group of subjects to be remarkably accurate.

The possibility should be considered, however, that this apparent accuracy might actually have resulted from the cancellation of two or more systematic errors acting to bias the results in opposite directions. For example, it might be that the rotating spot method itself produces results that are systematically in error by about 125 ms, because the neural processing that culminates in a visual object's finally entering consciousness takes about 125 ms. This might mean that, for example, by the time the spot entered the subject's awareness at clock position 2, it would actually have already moved on to clock position 4 or 5—so that the subject would systematically report events as occurring before they actually did. This scenario would certainly account for Libet's original report that the average subjective time M was approximately -125 ms in relation to the button press. Libet himself discounts this possibility on the grounds that he had already shown that sensations take not 100 ms but as much as 500 ms to enter awareness, but are then ''subjectively back referred'' to the time of the external stimulus (Libet et al., 1983). However this argument ignores the fact that the interpretation of the experiments Libet takes to demonstrate the existence of subjective back referral is very controversial (e.g., Churchland, 1981a, 1981b; Glynn, 1990; Gomes, 1998, 2002; Pockett, 2002a, 2002b).

In any case, if the rotating spot method does produce systematically biassed judgements because of the above considerations, these errors must in our case have been cancelled out by one or more further systematic errors of either perception or reporting that act in the opposite direction. Possible candidates include the effects reported by:

(a) Joordens, Van Duijn, and Spalek (2002), who show that when subjects are asked to use the clock method to report the time they perceive the clock face to change colour, they remember the colour change as occurring 70 ms later than it actually did. The authors speculate that this effect could be due either to a memory bias or a decision bias. It is not entirely clear that the subjective perception of an externally generated event like a colour change can legitimately be compared with perception of an internally generated event like a decision to move, but if it can, this finding suggests that our subjects may have perceived their decision times as being later than they actually were (which would tend to cancel the effect produced by sensory delay).

(b) Haggard and colleagues (for example Haggard & Clark, 2003), who have investigated what they call the intentional binding effect. This effect dictates that the times of intended actions are perceived as being closer to the effects of the action than they really are, while unintended actions are perceived as more separated from the effects than they really are. The intended actions in our experiments have the effect of stopping the rotation of the spot, so it is possible that the intentional binding effect might lead the subjects to perceive their key press as being later than it really was (which would also tend to cancel the effects of sensory delay).

Of course, in order to postulate that this sort of cancellation did occur in our experiments, one would also have to postulate that it did not occur in Libet's original experiments. For the latter postulate to be true, there would have to have been additional errors introduced by Libet's protocol which were not present in ours. Possible sources of such error are discussed below.

4.3. Non-effects of other tested factors

Another major finding of the present study was that of the seven factors tested, six had no significant effect on the trimmed means of the data.

The non-effect of Urge/Decision (act spontaneously or decide NOW) is not unexpected, since there is no particular reason to suppose that there would be any difference in perceived time of moving (Libet's time M) between movements which were allowed to occur spontaneously and movements which resulted from a definite decision. However, our data here are useful in that they provide a suitable baseline for future investigations of whether a spontaneous urge is different from a definite decision when measuring Libet's time W.

The non-effect of Dark/Light Spot (dark blue spot vs lime green spot) is perhaps of more significance in terms of the accuracy of the method as a whole. This factor was investigated because both Klein (2002) and van de Grind (2002) have suggested that data produced by the rotating spot method might be biased by the flash lag effect (Nijhawan, 1994). The flash lag effect causes misperception of the position of moving spots of light with respect to stationary flashed spots of light. The effect originally reported was that flashed lights appear to lag behind moving lights, but it was later demonstrated that variations in relative brightness of the flashed and moving lights could cause the flashed light either to lag or lead the moving lights (Lappe & Krekelberg, 1998; Purushothaman et al., 1998). Thus it is reasonable to predict that if the flash-lag effect were involved in biasing the results of the rotating spot method, varying the brightness of the spot in the rotating spot method would have at least some effect on the results obtained using that method. We were not able technically to control the brightness of our rotating spot independently of the brightness of the rest of the clock, but the two levels of factor 3 (lime green spot or dark blue spot on a black background) did serve to provide a subjectively obvious difference in ease of visibility of the rotating spot. The fact that there was no final statistical difference in our data between the results produced with lime green (easily visible) spots and those produced with dark blue (less easily visible) spots thus suggests that the flash lag effect is probably not materially involved in the outcome of experiments using the rotating spot method. It should be noted that factor 3 *was* statistically significant before the adjustment for multiple tests, which suggests that either a very much larger experiment or a very much greater luminance difference between the two levels of the factor could conceivably result in a result that ended up being statistically significant. But even then, whether such an outcome would validate the hypothesis that the flash-lag effect affects the rotating spot method is doubtful, since a spot with extremely low visibility could be expected to alter results on any number of different grounds. We conclude that within reasonable bounds, the flash lag effect does not produce any systematic bias in the results of the rotating spot method.

Centre/Follow (whether the subjects fixed their gaze on the centre of the clock or followed the spot round with their eyes) was significant before the adjustment for multiple tests, but not after. Table 4 shows that inasmuch as there was any effect, the results tended to be more negative when the subject followed the spot around with their eyes than when they fixed their gaze on the centre of the circle.

Other investigated factors which turned out to have no effect on the trimmed means were the diameter of the clock, the diameter of the rotating spot and the speed of rotation. These findings make it unlikely that variations in any of these factors would cause differences in the results obtained in different laboratories.

Overall the above findings suggest that the rotating spot method is remarkably robust, in that minor differences in any of the parameters studied (except for Start/End Movt) would not affect the outcome. Since Start/End Movt was the only factor that *a priori* should have affected the outcome, our results can be considered a strong validation of the rotating spot method for the timing of at least this kind of subjective event.

4.4. Factor settings that most reduce variability of results

Even though most of the factors studied did not influence the trimmed means of our data, the experiments did throw up a certain combination of factors which produced the least within-subject variability. This combination was making a definite decision to ACT NOW (rather than allowing the urge to act to arise spontaneously) and using a small circle with a fast rotation speed. In some senses, the success of this combination in reducing intra-subject variability is counterintuitive, since a number of subjects volunteered the observation that the fast rotation/small circle option presented the most difficult task in the whole experiment. It may be that the requirement to concentrate harder in order to do this more difficult task was beneficial from the point of view of reducing variability. As to why the requirement to report the instant of an action resulting from a definite decision as opposed to a mere urge produced less variable results, it may similarly be that internally saying the word NOW helped to focus attention on the task during the time period in which the movement happened.

4.5. Possible reasons for differences in previously reported mean times of Libet's M

There are a number of possible reasons for the differences between our results (−4 ms), Libet's original results (−125 ms) and Haggard's original results (−89 ms) with regard to reported "time M"s:

(a) Use of simple means rather than trimmed means or medians in the earlier reports may have introduced bias due to the presence of outliers.
(b) Use of very long recording sessions in the earlier reports may have introduced bias due to increasing tiredness or lack of interest in the task. (Our data provide no evidence to support this suggestion, but we deliberately kept our sessions very short with the aim of obviating such an effect).
(c) None of the studies presently in the literature specifies in their instructions to subjects whether the subjects should report the time at which their movements started or the time at which they finished. As shown by our data, different proportions of subjects deciding to report one or the other time in different experiments could potentially produce differences of up to 40 ms in average subjective reports.
(d) The relatively small samples of subjects used in the earlier reports may have introduced a significant sampling error. Both our data and those of Libet et al. (1983) demonstrate major differences between individuals. Small samples could easily include by chance a greater percentage of any particular sort of individual than is present in the population under study.
(e) Independent of sampling error, the actual populations sampled in the different reports in the literature may have been intrinsically different, with regard to either their perceptions or their ability to report those perceptions. Our population was composed almost entirely of university physicists, who may be either constitutionally or as a result of training more capable of accurate observation and reporting than the populations studied in earlier reports.
(f) Perhaps the single most important difference between our protocol and that of Libet et al. (1983) is that Libet gave all of his subjects an large amount of pretraining, while almost all of our subjects were completely naïve. Libet et al. (1983) report that "The first (and in some cases the second) half-day session was purely for training purposes. Subsequently, each subject was studied in 6–8 regular half-day sessions, usually 1 per week. Each of the first 4 regular sessions began with a training series of 25 trials [with feedback] ... [and each of the following series of trials in that session was also] preceded by a briefer 10-trial series for retraining purposes." Libet was measuring times W and S as well as time M, but the fact remains that his protocol contrasts markedly with our single 20-min session for each subject, which for most subjects included no pretraining at all. It seems to us highly likely that feedback from the experimenter may have been a major factor acting to introduce a difference between Libet's results and ours.

4.6. Recommendations for future experiments using the rotating spot method

If an "accurate" outcome (in the sense that the subjective estimate of movement time over all subjects coincides well with the objectively measured movement time) and minimum variability in individual results are desiderata, it seems from our results that the best combination of experimental factors and analysis methods for future investigations would involve:

- naïve subjects
- short recording sessions
- small diameter clock and fast rotating spot
- precise instructions to the subjects—for example if the subjective event to be reported is the perceived time of a voluntary movement, subjects should be told whether to pay attention to the start or the end of the movement
- analysis of medians or trimmed means rather than simple means
- subject numbers as large as possible.

When these recommendations are followed, the method produces robust results, which suggest that our awareness of the time at which we move is remarkably accurate.

Acknowledgments

The authors would like to thank Andrew Hay for writing the software and Associate Professor G.E.J. Bold for his general support and enthusiasm.

References

Churchland, P. S. (1981a). On the alleged backwards referral of experiences and its relevance to the mind-body problem. *Philosophy of Science, 48*, 165–181.

Churchland, P. S. (1981b). The timing of sensations: reply to Libet. *Philosophy of Science, 48*, 492–497.

Glynn, I. M. (1990). Consciousness and time. *Nature, 348*, 477–479.

Gomes, G. (1998). The timing of conscious experience: a critical review and reinterpretation of Libet's research. *Consciousness and Cognition, 7*, 559–595.

Gomes, G. (2002). Problems in the timing of conscious experience. *Consciousness and Cognition, 11*, 191–197.

Haggard, P., & Clark, S. (2003). Intentional action: conscious experience and neural prediction. *Consciousness and Cognition, 12*, 695–707.

Haggard, P., & Eimer, M. (1999). On the relation between brain potentials and the awareness of voluntary movements. *Experimental Brain Research, 126*, 128–133.

Joordens, S., Van Duijn, M., & Spalek, T. M. (2002). When timing the mind one should also mind the timing: bias in the measurement of voluntary actions. *Consciousness and Cognition, 11*, 231–240.

Klein, S. A. (2002). Libet's temporal anomalies: a reassessment of the data. *Consciousness and Cognition, 11*, 198–214.

Lappe, M., & Krekelberg, B. (1998). The position of moving objects. *Perception, 27*, 1433–1437.

Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): the unconscious initiation of a freely voluntary act. *Brain, 106*, 623–642.

Mele, A. F. (2006). Free will: theories, analysis and data. In S. Pockett, W. P. Banks, & S. Gallagher (Eds.), *Does consciousness cause behavior?* (pp. 187–206). Cambridge, Mass.: MIT Press.

Nijhawan, R. (1994). Motion extrapolation in catching. *Nature, 370*, 256–257.

Pockett, S. (2002a). On subjective back-referral and how long it takes to become conscious of a stimulus: a reinterpretation of Libet's data. *Consciousness and Cognition, 11*, 144–161.

Pockett, S. (2002b). Backward referral, flash-lags and quantum free will: a response to commentaries on articles by Pockett, Klein, Gomes and Trevena and Miller. *Consciousness and Cognition, 11*, 314–325.

Purushothaman, G., Patel, S. S., Bedell, H. E., & Ogmen, H. (1998). Moving ahead through differential visual latency. *Nature, 396*, 424.

Sirigu, A., Daprati, E., Ciancia, S., Giraux, P., Nighoghossian, N., Posada, A., et al. (2004). Altered awareness of voluntary action after damage to the parietal cortex. *Nature Neuroscience, 7*, 80–84.

Spence, C., Shore, D. I., & Klein, R. M. (2001). Multisensory prior entry. *Journal of Experimental Psychology: General, 130*, 799–832.

Tsakiris, M., & Haggard, P. (2003). Awareness of somatic events associated with a voluntary action. *Experimental Brain Research, 149*, 439–446.

van de Grind, W. (2002). Physical, neural, and mental timing. *Consciousness and Cognition, 11*, 241–264.

Wohlschlager, A., Haggard, P., Gesierich, B., & Prinz, W. (2003). The perceived onset time of self- and other-generated actions. *Psychological Science, 14*, 586–591.

Wu, C. F. J., & Hamada, M. (2000). *Experiments: Planning, analysis, and parameter design optimization.* New York: Wiley-Interscience.