Available online at www.sciencedirect.com
SCIENCE @ DIRECT®
ELSEVIER
Vision Research 46 (2006) 1327–1330
Vision Research
www.elsevier.com/locate/visres

Lapse resistance in the verbal letter reporting task

Aries Arditi *

Arlene R. Gordon Research Institute, Lighthouse International, 111 East 59th Street, New York, NY 10022, USA

Received 11 March 2005; received in revised form 9 June 2005

Abstract

Lapses, or misreporting errors, can affect accuracy of threshold measurements. Assumptions about lapse rate, especially in untrained observers, have consequently guided the design of at least one clinical psychophysical test. Lapse rate was assessed using a verbal letter identification paradigm like that used in visual acuity and letter contrast sensitivity testing. Subjects occasionally made slip-of-the tongue errors but spontaneously corrected them. Lapse rate (excluding such errors) was 0–3 errors per 1536 (average rate of 0.0005). In this common clinical paradigm, in which observers set their reporting pace, and where opportunity to amend responses is available, lapse rate is negligible.

Keywords: Lapse rate; Misreporting rate; Extraneous noise; Clinical psychophysics; Acuity; Contrast sensitivity; Psychophysical testing; Psychometric function

1. Introduction

Lapses, also called misreports, or extraneous noise are errors that are unrelated to failures of stimulus detection or discrimination. They can occur even when an observer detects, discriminates, and identifies the stimulus perfectly. They may occur because the observer fails to pay attention to the task, because they are attending to the wrong stimulus, or in the case of automated psychophysical testing, because they have blinked, or have inadvertently pressed the wrong response button.

Lapse rate is usually modeled as λ, the departure from perfect performance in the upper asymptotic region of the psychometric function (ψ; see Fig. 1). Lapses are most evident at the upper asymptote of ψ, where they prevent perfect performance, but they may occur at any stimulus intensity. When they occur well below threshold in the lower asymptotic region of ψ, they have the same effect as chance responding and are thus moot. When they occur in the region of ψ where the slope is positive, they are an additional source of performance variance.

Because lapses are by definition unrelated to sensory aspects of performance, they are of limited substantive interest to psychophysicists, but because they can affect fitting of psychometric function data, their potential effects must be addressed. Failure to account for them accurately can result in biases and decreased precision of slope and location (threshold) parameters of ψ (Swanson & Birch, 1992; Wichmann & Hill, 2001). Since obtaining these parameter values, especially threshold, is often the sole purpose for collecting psychometric function data in the first place, it is important either to use models that are not sensitive to lapses, or to have a means to determine, with reasonable confidence, what the lapse rate is.

Lapses are, to most experimenters, a nuisance. Because they occur infrequently, it is difficult to independently accurately estimate their rate. They may be estimated by adding lapse rate as another free parameter in a performance model, but to do so greatly increases the amount of data that must collected, for a given

* Tel.: +1 212 821 9500; fax: +1 212 751 9667.
E-mail address: aarditi@lighthouse.org.

0042-6989/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2005.06.014

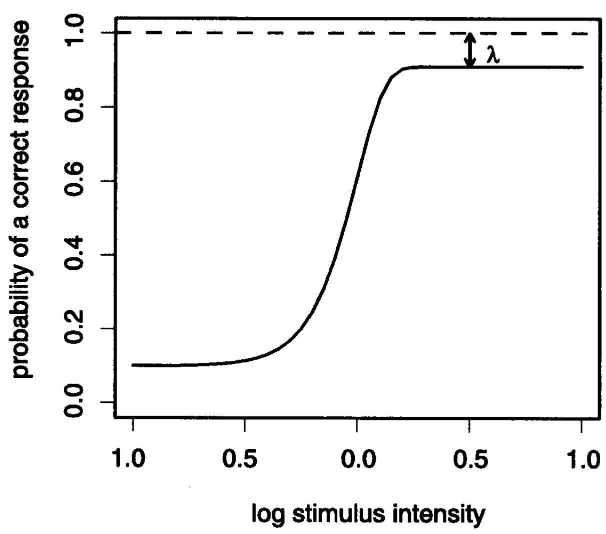

Fig. 1. The Weibull functional form of the psychometric function, showing λ, the lapse rate, as the departure from perfect performance, when stimulus intensity is substantially above threshold. In this example, the probability of a correct guess and the lapse rate are both 0.1.

goodness-of-fit (Klein, 2001). They are especially important to consider in the development of good clinical tests (Pelli, Robson, & Wilkins, 1988), which, for reasons of efficiency, must make assumptions about psychometric function parameters other than the one of interest (usually threshold), rather than estimate them from the data. Lapses are generally presumed to occur more frequently in clinical work than in basic psychophysics (Klein, 2001; Pelli et al., 1988), making it especially important to use paradigms in that setting that can be shown to have high accuracy even when substantial lapses are expected (Pelli et al., 1988) or that can be shown to have very low lapse rates (Swanson & Birch, 1992).

The purpose of this study is to assess frequency of lapses in the task of verbally reported letter identification as typically used in clinical visual acuity and contrast sensitivity measurements. Note that it is impossible to empirically estimate lapse rate at or near threshold, because in this region errors of detection or discrimination may be due *either* to sensory or extraneous factors, and there is no way to know which of these caused the errors. It is only in the upper asymptotic region where sensory factors can be ruled out, where the stimulus is unequivocally detectable, that empirical lapse estimates can be made. Thus, the measurements reported here all use unequivocally legible, decipherable, and suprathreshold stimuli.

2. Methods

Lapse rate was measured in 10 normally sighted, naïve observers, aged 22–82 years, using randomly generated, large, high contrast Sloan letters (Sloan, 1959) printed on white, letter size (8.5 × 11 in.) paper. Letters are used as optotypes in many clinical tests, including the Pelli–Robson Contrast Sensitivity Test (Pelli et al., 1988), the Mars Letter Contrast Sensitivity Test (Arditi, 2005), and the ETDRS acuity charts (Ferris, Kassoff, Bresnick, & Bailey, 1982). The use of large, high contrast letters made it reasonable to assume that errors observed were truly lapses and not related to problems in seeing or encoding the letters. The use of normally sighted observers made it reasonable to assume that errors were not due to a visual disorder.

Letters on each page were arranged in eight rows of six letters each, as on an acuity chart (except that the letters were all large—1.75 cm on a side, 2° at 50 cm, just discernible at that distance with acuity of 20/480), or a letter contrast sensitivity chart (except that the letters were all high contrast—close to 1.0). The letter sequence was random, except that, as in most letter test charts, the same letter never appeared in adjacent positions either vertically or horizontally. The first page of the 32-page sequence is shown in Fig. 2.

No practice was given, and the only experience subjects had in the specific task of reading large high contrast letters was presumably through prior routine eye care, screenings, and driver's licensure.

Participants were simply asked to read the letters aloud while holding the pages at a comfortable distance or on the table before them. The simple instruction to

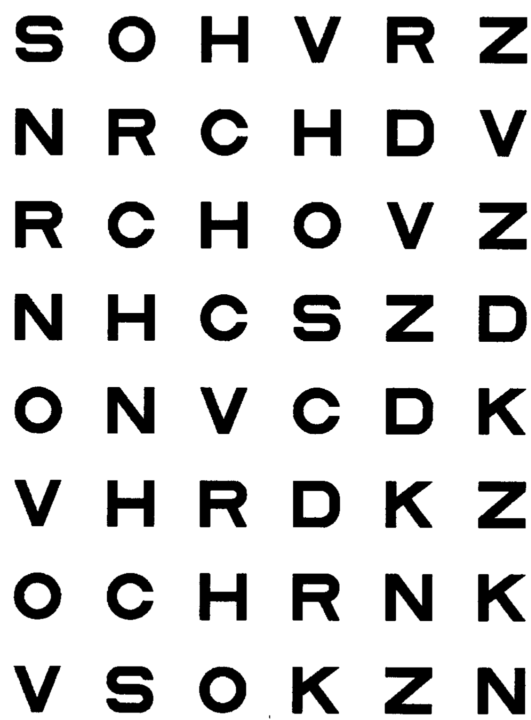

Fig. 2. Sample (first page) of the sequence of letters used in the experiment. Actual size of the page was 8.5 × 11 in.

read letters aloud is similar to a clinical testing situation. Each observer read 1536 letters.

The experimenter listened to the responses vigilantly and recorded errors, which were all assumed to be lapses, by hand on a clipboard. Because the letters "C" and "Z" are phonetically similar, there were a few instances in which the experimenter was unsure of which of these two letters was spoken. In these cases, the experimenter accepted the ambiguous letter as correct, and asked the participant to try to speak these letters clearly so as to make them easier to distinguish.

If an error was made and was not spontaneously corrected by the subject prior to the reading of the next line, the experimenter recorded it as an *initial* lapse. If an error was corrected prior to the reading of the next line, it was not counted as a lapse.

After finishing reading that page, the subject was asked to re-read the line with the lapse. Re-reading was delayed until the end of the page to avoid alerting the subject to the presence of an error. Errors on re-reading were recorded as *final* lapses.

3. Results

Results are shown in Table 1. Of a total of 15,360 letters read, only seven initial lapses were made, yielding an average initial lapse rate of 0.000456. Six of the 10 participants made no errors at all; the maximum initial lapse rate observed in a single participant was 0.002. The final (after re-reading) lapse rate was 0.0000. Thus the last column of Table 1 is all zeros, as would be expected from the low lapse rate.

4. Discussion

The average initial lapse rate result reported here is inconsistent with earlier ideas about lapses in clinical letter identification tests. Pelli et al. (1988), for example, presumed that lapses would be substantial in letter contrast sensitivity. They designed their chart using the same Sloan letters used in the present study, and based it on the predictions of a simple Weibull model. In applying their model to arrive at their final chart design, they assumed that λ was substantial, about 1% in experienced, and up to 5% in inexperienced, observers. Elliott, Bullimore, and Bailey (1991) assessed letter contrast sensitivity test–retest reliability empirically, and based on their findings and the model of Pelli et al. (1988) inferred that lapse rate was above 2%. Since, however, their estimate was based solely on *error rate*, which includes sensory, and lapse errors, this inference is not valid. Neither Pelli et al. nor Elliott et al. made empirical measurements of lapse rate. The present results suggest that lapses may be two or more orders of magnitude less frequent than these earlier estimates.

There are good reasons to believe that lapse rate in typical clinical test situations is probably substantially lower even than the 0.0005 observed here as the initial lapse rate. First, scoring differs from that of the present experiment in that in most situations it is common, and considered perfectly appropriate, to ask for the re-reading of a line where an error has been made, precisely so that the clinician can be sure that the error was visual in nature. Psychophysical purists, however, would argue that asking for re-reading of lines is a potential source of bias.

The present experiment simulates that situation by asking later, after the page is finished, for the re-reading of lines where errors were made. (Re-reading was not solicited immediately after the error was made, to avoid cueing the participant as to the location of a lapse.) In every case where an initial error was made, the error was corrected when the subject was asked to re-read the line with the error, so that the average final lapse rate was nil. In typical clinical test situations, then, where such re-reading is common, the effective lapse rate is probably much lower even than the 0.0005 average initial error rate observed here.

A second reason that lapse rates might be much lower is that in a procedure like the method of limits, lapse rate arguably becomes substantially lower as the observer approaches threshold. For example, in a visual acuity test, as letters become smaller and smaller, the observer is likely to attend more, scrutinize the stimulus, and take more time to respond than he or she might on the higher lines where the identifications are easy.

It is interesting to note that during the present experiment participants (especially those who read quickly) stumbled several times during the test, making errors when letters had phonetic similarities (e.g., "C" and "Z," "D" and "V"), or occasionally reversing the order of a pair of letters. Participants almost always revoked the incorrect responses and spontaneously corrected

Table 1
Ages, log MAR visual acuities, and lapses for the 10 participants of the study

Participant	Age	Log MAR	Initial lapses	Final lapses
DC	22	−0.08	0	0
PO	25	−0.08	0	0
JL	26	0.02	0	0
KB	33	0.02	1	0
CD	34	−0.08	0	0
VC	35	−0.08	2	0
HG	37	−0.10	1	0
TD	62	0.02	0	0
EF	81	−0.10	3	0
MC	82	0.32	0	0

Initial lapses were those made and not spontaneously corrected during reading of each line. Final lapses were those that remained after later asking participants to re-read lines on which an initial lapse had been observed.

these errors before they reached the end of the line, so most of these transient errors did not have an impact on their performance in the experiment. A few, of course, were not spontaneously corrected, and were counted as lapses.

Such errors are probably analogous to many "button errors" we observe in automated experiments, except that button errors cannot be retracted and resubmitted, as they can in the present experiment. Average response rate ranged from 1.2 to 2.8 letters/s, but there are not enough data from the four participants who had nonzero lapse rates to draw any conclusions regarding the relationship between rate of responding and lapses.

What about lapse errors in the experimenter, who was recording and scoring the lapses of the participant? Might these have occurred, and if so, would not this result in an underestimate of the lapse rate? The experimenter might have lapsed in two ways: failing to record a lapse that occurred in the participant (which would lead to an underestimate of lapse rate), and lapsing by scoring a correct letter as incorrect (which would lead to an overestimate). Admittedly, since the overwhelming majority of responses are correct, the experimenter is subject to a response bias that makes the former kind of experimenter lapse more likely than the latter. It is difficult to know what the experimenter lapse rate might be from these data, though since the experimenter did not have to verbally produce letters, one might suppose it to be lower than that of the participants. However, even if the experimenter failed to record one lapse for every lapse observed, and all such lapses were of the kind that led to underestimating lapse rate, true lapse rate would still be only doubled, a small increase compared to the factor of 100 difference between the present results and earlier estimates.

Why have lapses become an issue in contemporary psychophysics? In earlier times, before psychophysics was conducted in rapid, automated experiments, responses were mostly verbal, and with the exception of reaction time experiments, almost always revocable. Nowadays, however, computer technology gives us the opportunity to collect copious amounts of threshold data at high rates, using "expert" observers who press response buttons at a furious pace and often make judgments within a single second. These observers are not given the opportunity to change their response, nor to take a second try at a perceptual judgment, the way they do with clinical vision testing, because the press of a button, keyboard key, or some other "ballistic" device does not allow it. Indeed, expert observers usually are instructed (or instruct themselves) that a small number of errors are to be expected as "button" or "blink" errors, and that they will make only a small difference when the data are averaged. From this perspective, the lapses of contemporary psychophysics are a cost in reduced accuracy that we trade for being able to collect huge amounts of threshold data.

5. Conclusion

The present results, contrary to earlier suggestions (Elliott et al., 1991; Pelli et al., 1988) indicate that the clinical psychophysical procedure of verbal reporting of letters may be inherently resistant to lapses. They also suggest that in this task, and perhaps in some other related psychophysical paradigms, lapses are extraordinarily infrequent, and can safely be ignored. Finally, they show that verbal reporting, and other procedures that encourage slower and revocable responses, may reduce or eliminate the need to include the nuisance parameter of lapse rate in psychometric function modeling of empirical data.

Acknowledgment

This work was supported by US National Institutes of Health Grant EY015192.

References

Arditi, A. (2005). Improving the design of the letter contrast sensitivity test. *Investigative Ophthalmology and Visual Science, 46*, 2225–2229.

Elliott, D. B., Bullimore, M. A., & Bailey, I. L. (1991). Improving the reliability of the Pelli–Robson contrast sensitivity test. *Clinical Vision Science, 6*, 471–475.

Ferris, F. L., Kassoff, A., Bresnick, G. H., & Bailey, I. L. (1982). New visual acuity charts for clinical research. *American Journal of Ophthalmology, 94*, 92–96.

Klein, S. (2001). Measuring, estimating, and understanding the psychometric function. *Perception and Psychophysics, 63*(8), 1421–1455.

Pelli, D. G., Robson, J. G., & Wilkins, A. J. (1988). The design of a new letter contrast chart for measuring contrast sensitivity. *Clinical Vision Sciences, 2*, 187–199.

Sloan, L. L. (1959). New test charts for the measurement of visual acuity at far and near distances. *American Journal of Ophthalmology, 48*, 807.

Swanson, W. H., & Birch, E. E. (1992). Extracting thresholds from noisy psychophysical data. *Perception and Psychophysics, 51*, 409–422.

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception and Psychophysics, 63*, 1293–1313.