

NOTES AND COMMENT

Additional rules for the transformed up–down method in psychophysics

LAWRENCE G. BROWN

Johns Hopkins University, Baltimore, Maryland

In a classic paper, Levitt (1971) described an adaptive procedure for estimating points on the psychometric function known as the *transformed up–down method*. Levitt discussed the assumptions of the method and presented a brief table with simple rules that converge to a few different points on the psychometric function. Levitt's original table contains only the simplest rules, and sparsely covers the range of the psychometric function. This paper provides a table with previously unpublished rules which cover the range of the psychometric function at 5% intervals. There is a brief review of the major issues in adaptive testing. Technical issues such as the mean length and logical construction of the new rules are discussed.

The most widely used adaptive procedure for measuring sensory thresholds is the up–down staircase method (Cornsweet, 1962; Dixon & Mood, 1948; Levitt, 1971; Wetheril & Levitt, 1965). In this method, a subject's responses from one or more trials are used to determine the stimulus level. The responses are recorded and classified as either an "up" response or a "down" response according to rules appropriate for converging to a given point on the psychometric function. If an up response is obtained, the test stimulus level is increased by a certain amount (the step size). If a down response is obtained, the stimulus level is decreased by one step. The step size is constant and equal for up and down steps (if the experimenter can relax the requirement of a constant step size, then a somewhat simpler, modified version of the up–down method is available; see Kaernbach, 1991).

The basic form of the transformed up–down method will converge on the stimulus level for which the proportion of positive responses is 50% (X_{50}) and is called simply the *up–down method*. A generalization of the method is the *transformed up–down method*, which is useful for estimating points other than X_{50} . The transformed up–down method was most clearly presented by Levitt (1971), who provided a table containing rules that converge to several different points on the psychometric function.

The popularity of the up–down staircase method arises primarily from its simplicity, because the up–down procedure is not the most efficient. The most efficient methods are the maximum-likelihood procedures, such as

Watson and Pelli's (1983) QUEST. The up–down staircase methods are simple because they make only the weakest assumption about the psychometric function (that it be monotonic increasing) and typically require response groupings that are no more than a few trials in length. In comparison, maximum-likelihood procedures assume a particular form of the psychometric function and base the threshold estimate on the results of all trials tested.

Most researchers use the up–down staircase method with two-interval forced-choice (2IFC) stimulus presentations (Kershaw, 1985). Researchers use 2IFC because it has been demonstrated to be an essentially unbiased (in the signal-detection-theory sense) procedure (Macmillan & Creelman, 1991). Until relatively recently, if one wanted to be sure of unbiased responding, forced-choice designs were the only option. However, two new methods have been described for unbiased adaptive testing in a yes–no paradigm. The first is the single-interval adjustment matrix procedure of Kaernbach (1990), which adjusts the stimulus in a staircase manner. The second procedure is a maximum-likelihood method developed by Green (1993).

While the yes–no design is, in general, suspect because the subject's criterion is uncontrolled, this may not always turn out to be an issue in practice. For example, Pierce and King-Smith (1992) compared the reliability of visual thresholds (for a spot on a background) in 19 normal observers determined by an adaptive procedure (QUEST) obtained from yes–no or 2IFC presentations. No significant difference in test–retest reliability was found between the yes–no and 2IFC thresholds. This does not show that the visual threshold was completely unaffected by the subject's criterion (further statistical analysis did reveal an effect of false positive rate), but rather that for this specific task the criterion effect was relatively small.

The author wishes to thank C. Douglas Creelman and Christian Kaernbach for their valuable remarks on the manuscript. The author's address is Psychology Department, Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218 (e-mail: lgb@trolland.psy.jhu.edu).

Table 1
Response Groupings for Transformed Up-Down Rules

Entry	Response Sequences		<i>p</i>	MGL*	Reference
	Up	Down			
1	-	+	0.500	1.00	Levitt (1971)
2	--+- or +- or --	-++ or -++ or ++	0.550	2.36	
3	+-- or -	+--+ or ++	0.597	1.84	
4	+++ or +- or --	+++ or -+	0.648	2.42	
5	+- or -	++	0.707	1.71	Levitt (1971)
6	+++- or ++-- or +- or -	++++ or +++	0.749	2.56	
7	++- or +- or -	+++	0.794	2.42	Levitt (1971)
8	++++ or +++ or +- or -	++++	0.841	3.14	Levitt (1971)
9	+++++- or ++++- or +++- or ++- or +- or -	+++++	0.891	4.58	Wetherill & Levitt (1965)

Note—The stimulus level should be increased after an up sequence and decreased after a down sequence.
*Mean group length.

Despite these recent advances in unbiased adaptive procedures, it is safe to say that the up-down method with 2IFC presentation is among the most widely used adaptive procedures for determining sensory thresholds. In 2IFC experiments, the observed performance ranges between 50% correct and 100% correct. Unfortunately, Levitt's original papers offer only four rules that converge to points in this range (X_{71} , X_{79} , X_{84} , and X_{89}). This limited coverage seems to have misled some researchers into thinking that no other rules are available. For example, Kaernbach (1991) states, "Unfortunately, there exists no transformed up-down rule for X_{75} ."

Two points should be made about this statement. First, it seems unlikely that most researchers believe that Levitt's rules were the only rules possible. A more reasonable interpretation of Kaernbach's statement is that experimenters were lulled into believing that to achieve other convergence points with the transformed up-down method would be prohibitively complicated. The major point of this paper is to dispell that myth. Levitt published a rather restricted set of rules, and, without adding extreme complexity, many more rules are available. Table 1 presents rules that converge to points on the psychometric function (including X_{75}) at 5% intervals. Specific discussion of the new rules and their properties is in the next section.

The second point concerns the definition of a "threshold" value on the psychometric function. As Kaernbach

(1991) describes, many researchers who use 2IFC choose to define X_{75} as the threshold because it is the halfway point. Unfortunately, X_{75} is not necessarily the optimal point on the psychometric function for determining threshold. The optimal point is the one that minimizes the variance in the threshold estimate, called the *sweet-point* (Green, 1990; see also Taylor & Creelman, 1967). The sweetpoint has been shown to be in the X_{80} to X_{94} range, noticeably higher than the X_{71} value often used to approximate the halfway point. Unless the experimenter has a special reason for choosing X_{75} (to match earlier data, for example), it is not the optimal operational definition of threshold.

New Transformed Up-Down Rules

The transformed up-down method converges to the stimulus level where the probability of obtaining an up sequence equals the probability of a down sequence (or the probability of either equals .5). To solve for the point of convergence for a given rule, let *p* stand for the probability of a positive response at convergence and set the probability of an up sequence equal to the probability of a down sequence. For example, the simplest rule (Rule 1 in Table 1) involves stepping up for every negative response and down for every positive response. We therefore write $p = 1 - p$, which yields $p = .5$.

As a more complicated example, consider Rule 3. Setting $P(\text{up}) = P(\text{down})$ we write,

$$p(1-p)^2 + 1-p = p^2 + p^2(1-p).$$

The left side of this equation is simply the probability of obtaining an up response at convergence. The $p(1-p)^2$ term represents “+--” and $1-p$ represents “-”. The right side of the equation is written down analogously. This equation reduces to

$$2p^3 - 4p^2 + 1 = 0.$$

This equation can be solved numerically to yield $p = .59695$, recorded as $p = .597$ in Table 1. It is trivial to verify all the convergence points presented in Table 1 in the same manner.

The new transformed up-down rules were generated by a combination of two approaches. The first approach was to simply enumerate all the possible rules involving response sequences of 3 or less and calculate the probability at convergence. This is not as difficult as it might seem. There are only four response patterns of length 2 and eight patterns of length 3. Furthermore, only sequences with a negative response at the end make sense as an up sequence and vice versa. Some rules are quickly seen to be analogues of the simple up-down rule and must converge to X_{50} . These restrictions yield only about two dozen rules to solve if sequences no longer than 3 are allowed.

The rules involving longer response sequences (Rules 2 and 6) were discovered by a trial-and-error process guided by some observations of how target probabilities change by allowing longer sequences. For example, notice how Rules 1, 5, 7, 8, and 9 form a pattern; there is only one pattern in the down group, and the down pattern for Rule 5 is just the down pattern for Rule 1 with one more plus, and similarly for Rules 7 through 9. The target value of p for these rules is easily calculated. We have simply $p^n = .5$, where n is the number of “+” signs in the down pattern. Thus, for Rule 8 we have $p^4 = .5$, or $p = .84$. For Rule 9, $p^6 = .5$ or $p = .891$. Thus, one can easily create rules that converge to values of p arbitrarily close to 1 by simply using rules of the same form as 7 through 9, but with longer sequences.

Here is another example of how allowing longer patterns permits more points on the psychometric function to be achieved. Let us start with the following rule. Step up if one of the following patterns is obtained, -, +-, or +-+-. Step down if one of the following is obtained: +++ or +-+-+ (assume we have found this rule by exhaustive listing of patterns with 4 responses or less or simply by trial and error). As usual, we set $P(\text{up}) = P(\text{down})$ at convergence and write,

$$1 - p + p(1 - p) + p^2(1 - p)^2 = p^3 + p^3(1 - p),$$

which reduces to $2p^4 - 4p^3 + 1 = 0$, the solution of which is $p = .733615$. Thus, this rule converges to X_{73} (this rule is mentioned in Wetherill & Levitt, 1965). We can modify this rule slightly by allowing for patterns of length 5, and have a rule that will converge to X_{75} .

We do this by observing that if we have a rule for X_{73} and we want X_{75} , we need to step down less often. We

can do this by “subdividing” the pattern +-+-. We subdivide this pattern into two patterns by adding another response at the end. If we add a + at the end, we have another down pattern, and if we add a - at the end, we have an up pattern. What we have done is to subdivide the pattern +-+-+ of the rule for X_{73} into the two patterns +-+-- and +-+-- to make a new rule. This new rule is shown as Rule 6 in Table 1. With this new rule, we are going to step down only about half the time we get the pattern +-+-+ (that is, on those occasions when we get the pattern +-+-+ instead of every time, as in the rule for X_{73}). Solving the resulting polynomial of degree 5 yields $p = .749455$.

Of course, there are other rules for other points on the psychometric function. By simply exhausting the possible rules involving patterns of 4 responses or less, one can find rules that converge to X_{64} and X_{58} . Table 1 lists rules that cover the range of p from .5 to .9 in steps of .05, which should be sufficient for most purposes. Note that each of the rules in Table 1 is really two rules. Taking the “mirror image” of a rule that converges to p yields a rule that converges to $1-p$. To get the mirror image of a rule, make every + a -, and vice versa.

It is generally true that Levitt listed the points on the psychometric function with the *simplest* rules. While the new transformed up-down rules do not use any more patterns than the rules appearing in the original papers (the rule for X_{75} in Table 1 uses only six response patterns, four up and two down), the new rules are more complicated in other ways. For example, one may easily describe the original rules verbally in the following way: Step down after N successive correct responses and step up immediately after any incorrect response. In Table 1, this description fits Rules 1, 5, 7, 8, and 9, with N ranging from 1 through 6. Clearly, no such simple description fits Rules 2 or 6.

The complexity of a rule can be quantified by calculating the mean group length (MGL) for that rule. A “group” is either an up or a down pattern. MGL is calculated by multiplying the length of a group by the probability of obtaining that group (at convergence) for each group in a rule, and summing over all groups. For example, the MGL for the well-known X_{71} rule is $1(.293) + 2(.293)(.707) + 2(.707)(.707) = 1.71$. The MGL for each rule is shown in Table 1. If the staircase termination is specified by a fixed number of reversals, rules with a larger MGL should be slower to terminate, and the MGL for X_{75} is 50% longer than the MGL for X_{71} . The other new rules have shorter MGLs.

Finally, fewer logical constraints were enforced when the new rules were generated. In the original rules a - response always implied an immediate step up, but this is not the case for the new rules, which sometimes require two consecutive - responses. The only logical requirement is that up rules end on a - and down rules end on a +. This loosening of theoretical constraints allows for the possibility of rules like number 4 for X_{65} , where a distinction is made between +- (an up response) and -+ (a down response).

Conclusions

Table 1 should prove useful to researchers using the transformed up-down method for threshold estimation. The new rules allow much finer sampling of the psychometric function, without adding significant complexity. It is not claimed that the transformed up-down method is always the best adaptive procedure. The pros and cons of various alternatives were discussed in the introduction. However, the up-down method is easy to implement and reasonably efficient (Kershaw, 1985; Rose, Teller, & Rendleman, 1970). Certainly, the method should not be discarded because of inadequate sampling of the range of the psychometric function.

REFERENCES

- CORNSEWET, T. N. (1962). The staircase-method in psychophysics. *American Journal of Psychology*, **75**, 485-491.
- DIXON, W. J., & MOOD, A. M. (1948). A method for obtaining and analyzing sensitivity data. *Journal of the American Statistical Association*, **43**, 109-126.
- GREEN, D. M. (1990). Stimulus selection in adaptive psychophysical procedures. *Journal of the Acoustical Society of America*, **87**, 2662-2674.
- GREEN, D. M. (1993). A maximum-likelihood method for estimating thresholds in a yes-no task. *Journal of the Acoustical Society of America*, **93**, 2096-2105.
- KAERNBACH, C. (1990). A single-interval adjustment-matrix (SIAM) for unbiased adaptive testing. *Journal of the Acoustical Society of America*, **88**, 2645-2655.
- KAERNBACH, C. (1991). Simple adaptive testing with the weighted up-down method. *Perception & Psychophysics*, **49**, 227-229.
- KERSHAW, C. D. (1985). Statistical properties of staircase estimates from two interval forced choice experiments. *British Journal of Mathematical & Statistical Psychology*, **38**, 35-43.
- LEVITT, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, **49**, 467-477.
- MACMILLAN, N. A., & CREELMAN, C. D. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.
- PIERCE, G. E., & KING-SMITH, P. E. (1992). Yes-no or two alternative forced choice: Which is the better clinical threshold technique? *Investigative Ophthalmology & Visual Science* [Abstract], **33**(Suppl.), 964.
- ROSE, R. M., TELLER, D. Y., & RENDLEMAN, P. (1970). Statistical properties of staircase estimates. *Perception & Psychophysics*, **8**, 199-204.
- TAYLOR, M. M., & CREELMAN, C. D. (1967). PEST: Efficient estimates on probability functions. *Journal of the Acoustical Society of America*, **41**, 782-787.
- WATSON, A. B., & PELLI, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, **33**, 113-120.
- WETHERILL, G. B., & LEVITT, H. (1965). Sequential estimation of points on a psychometric function. *British Journal of Mathematical & Statistical Psychology*, **18**, 1-10.

(Manuscript received May 15, 1995;

revision accepted for publication November 26, 1995.)

Call for Nominations Troland Research Awards

The National Academy of Sciences requests nominations for the 1997 Troland Research Awards, awarded annually to young investigators to recognize unusual achievement and to further empirical research in psychology regarding the relationships of consciousness and the physical world.

The awards were established by the bequest of Leonard T. Troland, author of *The Principles of Psychophysiology*. Each of two awards carries with it a sum of \$35,000 to help further the awardee's research. Please note that recipients must be 40 years of age or younger and that the selection committee will give preference to empirical, quantitative research.

Nominations for these awards are due September 1, 1996. For further information or to obtain a nomination form, contact:

Ms. Rina Mathless
National Academy of Sciences
2101 Constitution Avenue, NW, Room 185
Washington, DC 20418
Phone: (202) 334-2444
Fax: (202) 334-2153
Internet: nasmembr@nas.edu
WWW: <http://www.nas.edu/nas>