THE CONTROL OF ATTITUDE IN PSYCHO- PHYSICAL EXPERIMENTS

BY EDWIN G. BORING

Clark University

Dr. Godfrey Thomson has recently published some very illuminating discussions of the mathematical logic of psychophysics, and I for one feel myself too much in his debt to to indulge in the mere picking of flaws. Nevertheless his paper, 'A New Point of View in the Interpretation of Threshold Measurements in Psychophysics,'¹ gives me concern because it seems to cast aspersions upon what I have regarded as the most promising direction of development in psychophysics. It is true that Dr. Thomson's 'offences' are implied rather than actual and that ultimately I may discover him in agreement with my thesis; nevertheless his article furnishes a reason for placing the point of view that I have in mind squarely before psychologists. I may add that I conceive that I am merely explicating an idea that arose within the Cornell Laboratory. If there be any credit for its origin it is due Cornell, though the responsibility of the present exposition is mine.

Dr. Thomson's suggestion is as follows: In the test of Weber's law (he is thinking of lifted weights as the example) we may take the differential threshold, which is half the distance between the upper and lower limens (*i.e.*, half the 'interval of uncertainty'), as the measure of sensitivity; and such has been the usual practice. Dr. Thomson, however, prefers to take as a measure of sensitivity the interquartile range of the point of subjective equality. The threshold is proportional to the distance between the two points where the psychometric functions for 'greater' and for 'less' cross the 50 per cent. abscissa, and its amount

¹ Godfrey H. Thomson, 'A New Point of View in the Interpretation of Threshold Measurements in Psychophysics,' *PSYCHOL. REV.*, 1920, 27, 300-307.

depends upon the number of judgments that fall within the third category of 'undecided' judgments (as Dr. Thomson styles them) or 'equal' judgments. Dr. Thomson would take these 'undecided' judgments and divide them equally for every stimulus value between the 'greater' and 'less' categories, thus establishing two new psychometric functions (instead of the original three) which of necessity will intersect upon the 50 per cent. abscissa and give zero limens. The interquartile range is the distance between the two points of intersection of these new psychometric functions with the 75 per cent. abscissa. It is independent of the 'undecided' judgments and dependent upon the measure of precision (h) of the psychometric functions, since the steeper the curves the less the interquartile range, and *vice versa*.

Dr. Thomson's preference for the interquartile range as a measure of sensitivity lies in his distrust of the relative frequencies of the 'undecided' judgments upon which the threshold depends. The threshold, he writes, 'depends entirely upon the subject's readiness to give the answer *undecided*. It measures therefore rather a moral character than a physical sensitivity.' 'The moral character of the measure $S - S'$ is above all seen from the fact that any subject who wishes may reduce it to zero, whatever may be his actual sensitivity, simply by determining that he will never give an answer *undecided*.' Thus the interquartile range 'is more physiological than the threshold measure.' The threshold is to be mistrusted because 'the decision as to what proportion . . . is to be called heavier, what undecided and what lighter depends upon a conscious act of the subject, and can be varied, if he be so disposed, at his whim; and will vary with his mood at the moment.' It is such a designation of things constant and measurable as 'physiological' and of things inconstant and uncontrolled as 'moral' or psychological that moves a psychologist to reply.

It is not my present purpose to inquire how much of Dr. Thomson's argument is actually new. Fechner wished to measure *Unterschiedsempfindlichkeit* by h and divided his

zweifelhafte cases between the 'greater' and 'less' categories.¹ G. E. Müller opposed Fechner, arguing that the limen must be used as the basis of Weber's law.² But Fechner stuck to his guns,³ as did also Müller.⁴ In 1904 Müller was wondering 'wie in aller Welt kann man ohne weiteres voraussetzen' that the measure of precision, which is independent of many factors entering into the lifting of weights, could constitute the basis of Weber's law. Dr. Thomson may need to meet Müller's argument against Fechner.⁵ I belong to the younger generation to whom Weber's law and *Unterschiedsempfindlichkeit* are less sacred than they once were, and I am willing to admit that the matter may well wait for supporting facts. Experimental studies are wanted that show both the threshold and the interquartile range as functions of the absolute magnitude of stimulus, and then we can determine how each fits the Weber-Fechner formula. But I am not willing to let Dr. Thomson dissuade us in advance from an interest in the threshold because its 'morality' can not be controlled. There is enough experimental work to indicate, so it seems to me, that accurate control of the third psychophysical category is possible and scientifically necessary.

THE EXPERIMENTAL CONTROL OF THE PSYCHOPHYSICAL JUDGMENTS

1. In the first place, if we are to gain accuracy of definition of the psychophysical categories, we must exclude the doubtful judgments.

Dr. Thomson has ample historical ground for including them. They have been left in from the first. Fechner and Müller called them the 'z-cases' (*zweifelhaft*), as we have seen. Müller raised the question as to whether 'doubtful' and 'equal' judgments ought not to be separated, but concluded

¹ G. Th. Fechner, 'Elemente der Psychophysik,' 1889, I, 101ff.; 'Revision der Hauptpunkte der Psychophysik,' 1882, 24f.

² G. E. Müller, 'Zur Grundlegung der Psychophysik,' 1878, 28f., 33-36.

³ Fechner, 'Revision,' 48f.

⁴ Müller, 'Die Gesichtspunkte und die Tatsachen der Psychophysischen Methodik,' 1904, 104-109.

⁵ E. B. Titchener summarizes the controversy: *Experimental Psychology*, 1905, II, ii, 278-285.

on experimental evidence that the positive impressions of 'equal' were rare, especially in trained observers, and that separate treatment of them was therefore not necessary.¹ Titchener in 1905 grouped both 'doubtful' and 'equal' cases under the heading 'u' (uncertain) or '?'.² Urban in 1908 had his subjects guess when in doubt which weight of the lifted pair was the heavier. Thus he obtained the categories 'heavier-guess' (hg) and 'lighter-guess' (lg), which he saw fit later to combine as 'equality' judgments.³ In such a setting it is natural to call the region in which the 'equality' judgments are most frequent the 'interval of uncertainty.' Urban's pupils have followed his final practice. Fernberger wrote: "The equality judgment was more complex [than the other judgments] as it not only included cases of actual subjective equality . . ., but also all those cases where it was impossible for the subject to give either a lighter or a heavier judgment, usually termed doubtful cases."⁴

It is under this practice that Dr. Thomson would reject the interval of uncertainty as a measure of sensitivity because 'any subject who wishes may reduce it to zero . . . simply by determining that he will never give an answer *undecided*.' He is undoubtedly thinking of results like Fernberger's on the effect of attitude on the interval of uncertainty; but plainly he does not accept, as I think he should, Fernberger's contention that attitude can be controlled in the laboratory and that, in view of this state of affairs, it must be controlled.⁵

Fernberger's proposal was that attitude should be rendered constant by explicit instructions and observational training,

¹ Müller, 'Methodik,' 12f.

² Titchener, *op. cit.* II, i, 107ff; ii, 268.

³ F. M. Urban, 'The Application of Statistical Methods to the Problems of Psychophysics,' 1908, 5f., 15, 99f., 106f., 110f., etc.

⁴ S. W. Fernberger, 'On the Relation of the Methods of Just Perceptible Differences and Constant Stimuli,' *Psychol. Monog.*, 1913, No. 61, 16.

⁵ Fernberger, 'The Effect of Attitude of the Subject upon the Measure of Sensitivity,' *Amer. J. Psychol.*, 1914, 25, 538-543. Fernberger refers here to another case in his own experiments and to Warner Brown's results. He has since called my attention to a study where three of eight subjects give no interval of uncertainty, presumably because of failure of attitudinal control: A. L. Ide, 'The Influence of Temperature on the Formation of Judgments in Lifted Weight Experiments,' 1919, 25 pp. (Univ. Pennsylvania thesis).

but I am going further in insisting that the instructions and experimental setting must assure a practically complete elimination of doubtful judgments. This is the point of George's study,¹ which merits, I think, considerable attention from psychophysicists. George's thesis is that in a psychophysical experiment we are dealing with a series of mental states which are a continuous function of the given series of stimuli: we vary the stimuli and note the concomitant mental variation. Under the rules of scientific experiment everything else must be kept constant including the attitude and psychophysical constitution of the subject; if they are inconstant we can no longer tell of what our judgments are a function. Moreover, if we find that any particular category is in itself an indicator of a change in attitude, we must so arrange the experiment that this form of judgment will not occur or rule it out from the results if it does occur, since the information that it yields is beside the point of the psychophysical problem, which always proposes the establishment of the dependence of judgment upon varying stimulus. George finds doubt and the doubtful judgments to be the great offenders against constancy. His method is the establishment of an *Einstellung* for constancy of attitude and the determination of (1) what categories the maintenance of the *Einstellung* rules out, (2) what categories, conversely, are noted by the observers as interfering with the *Einstellung* when they do appear, and (3) the objective evidence of attitudinal shift furnished by reaction times. George's article must be its own summary; I can not do it justice here. It is scarcely more than preliminary and his method itself may yet be called in question. It would be a slender weapon with which to combat the main body of psychophysical practice were it not for the fact that, like Dr. Thomson, many psychophysicists already know that something or other must be done to gain greater constancy of results, that the case against the doubtful judgment is plausible *a priori* [is a doubtful difference *ipso facto* less than an undoubted one?],

¹ S. S. George, 'Attitude in Relation to the Psychophysical Judgment,' *Amer. J. Psychol.*, 1917, 28, 1-37.

and that the attitude that George proposes as an ideal has already gained some slight support from experimental usage.¹ My personal prejudice for George's conditions I shall mention presently.

2. Doubt is the most persistent offender and the one presumably most responsible for Dr. Thomson's strictures upon the interval of 'uncertainty,' but other attitudinal seducers must also be dismissed. George makes a case against expectation, a case which indicates among other things that a haphazard method of presentation is to be preferred to a serial method with partial knowledge (*e.g.*, the method of least perceptible differences).

'Reflective' judgments are the general class under which inconstancy is apt to occur. George does not find that the reflective attitude necessarily means uncontrolled variability but merely that it favors it. The reflective judgment is very often equivocally determined; the subject is in a dilemma as to his report and *decides* what to report. He may be judging on the basis of more than one criterion, and may find in a given case different categories concurrently indicated by different criteria. If he must make a univocal report for such an equivocal situation, he must make some decision, even though the decision may be for the category 'undecided.' It is the occurrence of such uncontrolled 'decisions,' I take it, that makes Dr. Thomson wish to give up the threshold as a measure of sensitivity; but his rejection of the threshold is not necessary on this score, since reflective judgments can be avoided.

George shows that doubtful judgments tend to be reflective. The doubting subject is resolving a dilemma in favor of one category or the other, and does justice to the unreported category by labelling his judgment 'doubtful.' The 'or-judgments' (*e.g.*, 'equal-or-less') are compromises in which neither side has won out. The category 'no-difference' is also often a reflective compromise. Reflection and hesitation are, however, not synonymous; a judgment may be long

¹ L. B. Hoisington, *Amer. J. Psychol.*, 1917, 28, 588ff.; M. Kincaid, *ibid.*, 1918, 29, 227-232; A. M. Bowman, *ibid.*, 1920, 31, 87-90; C. C. Pratt, *ibid.*, 1920, 31.

delayed and yet come as a simple report of mental process without any indication of resolution or of attitudinal inconstancy. In fact it seems probable that the time of formation of the judgment is one of the mental factors which under a constant attitude is a serial function of the stimulus and a feasible subject for exact psychophysical investigation.

The type of judgment furthered by ordinary psychophysical procedure is shown in the introspective analyses in Fernberger's recent monograph.¹ These descriptions he obtained under conditions analogous to Urban's except in so far as the introspection itself interfered. Compared with what I have in mind as the ideal, they show a relatively complex—often very complex—process of comparing. I should not on the basis of them expect even the degree of constancy which one actually does get. The equality judgments are typically reflective in that they involve a 'verification process,' which Fernberger, as if in support of George, seems to equate to 'doubt.'²

Of course what is needed to support my argument, and what is lacking, is the companion introspective study made under George's conditions. Perhaps these consciousnesses would not prove so simple as I think. My conviction that the two consciousnesses would, however, be very different affairs is dependent, I must in honesty confess, upon personal experience. I was one of George's observers and I have observed in other experiments under his conditions. I have also observed with lifted weights on the turning-top table under conditions patterned after Urban's. And the two consciousnesses are to me almost unbelievably different. Under the conventional procedure I am constantly forced into resolutions, verifications, decisions, like those that Fernberger describes, and thus I am led into doubt and discomfort, and thence into a naïve uncritical attitude which affords no assurance of rigorous constancy. Under George's conditions enough effort is required to be sure, but the mental process is kept simple or else rejected. I do not have to report a

¹ Fernberger, 'An Introspective Analysis of Comparing,' *Psychol. Monog.*, 1919, No. 117.

² E.g., *ibid.*, p. 160.

complex situation by an inadequately simple word, nor an equivocal setting by a univocal judgment. It seems reasonable, does it not, that a report under George's conditions should be worth more than a report under Urban's, because it leads to attitudinal constancy or the detection of inconstancy when it does occur.¹

3. There is as much to be gained by the establishment of definite serial criteria of judgment as by the elimination of unsuitable categories of judgment. In fact, these two remedies are obverse and reverse. One can not get rid of doubt, for instance, unless the judgments are based upon a univocal criterion. Two criteria that may conflict are fatal to constancy since their resolution is left to chance.

The 'stimulus error' is a term which among other things characterizes judgments where a definite mental criterion is not established but judgment is left to chance habits. The term is undoubtedly much maligned as there are cases where the direction of the attention to the stimulus is psychologically useful, as in the preliminary investigation of a new perceptual field. But in general the stimulus attitude means indefiniteness and instability of criterion, as George pointed out.² A recent series of studies on the various criteria that may underlie the judgments of cutaneous duality shows how fundamental to accurate psychophysical work an avoidance of the stimulus attitude is.³ In lifted weights Friedländer's

¹ In justice to Urban it must be said that I do not believe he would experiment now as he did in 1908. Cf., e.g., the 'Statistical Methods,' 1908 (*op. cit.*), with his 'Ueber einige Begriffe und Aufgaben der Psychophysik,' *Arch. f. d. ges. Psychol.*, 1913, 30, 113-152. It is a pity that he has not yet been able to return to experimental work since writing his systematic articles. Moreover, both Urban and Fernberger have been under a special disadvantage in that they were dealing with a peculiarly refractory material, lifted weights. Isolation of a homogeneous series of univocal mental correlates of the stimulus is very difficult in the lifting process, although it has been attempted: H. Friedländer, 'Die Wahrnehmung der Schwere,' *Zeitsch. f. Psychol.*, 1920, 83, 129-120.

² George, *op. cit.*, 35f.

³ E. J. Gates, 'The Determinations of the Limens of Single and Dual Impression by the Method of Constant Stimuli,' *Amer. J. Psychol.*, 1915, 26, 152-157; Titchener, 'Ethnological Tests of Sensation and Perception,' etc., *Proc. Amer. Philos. Soc.*, 1916, 55, 206-215; E. deLaski, 'Perceptive Forms below the Level of the Two-point Limen,' *Amer. J. Psychol.*, 1916, 27, 569-571; C. L. Friedline, 'Discrimination of Two Cutaneous Patterns below the Two-point Limen,' *ibid.*, 1918, 29, 400-419.

experiments, although they leave much of psychophysical accuracy to be desired, seem at least to show that different numerical results follow when attention is upon the weight and when it is upon the sensory aspects of the lifting.¹

The more general ground for the control of criteria is the one which Fernberger took in defense of the interval of uncertainty.² The subjects must be constantly and effectively *eingestellt*, and the test of an effective *Einstellung* lies in preliminary trials, the taking of introspections, and the observer's full and repeated characterizations of their attitude.

4. The need for the isolation of the single judgment within the series is perhaps worth especial mention since the matter has just been [implicitly] brought to fore by Fernberger's measurement of the effect of one member of a series upon a succeeding member.³ Under Urban's conditions with the turning-top table Fernberger found that the judgment 'lighter' of a pair of lifted weights tends to be succeeded by a judgment 'heavier,' and *vice versa*. 'Fixing' a series so that one kind or the other of sequences predominates produces as startling effects upon the form of the psychometric functions as anything that Dr. Thomson is complaining of. Plainly some sort of expectational or rhythmic effect is operative; attitude is not remaining constant. Fernberger's solution of the difficulty is to balance one sort of succession against the other, and trust that they will cancel. For myself, I could not feel secure in such a procedure; an algebraic cancellation where the factors are so little understood can not be so satisfactory as an actual elimination. The members of the series should be separated by an interval—a distracted interval if necessary—so that the intraserial effects are broken up. This course slows down the rate of experimentation and robs the turning-top table of much of its charm; but, even when relative frequencies are aimed at, I do not conceive that numbers of cases can be allowed to weigh against rigorous scientific control.

¹ Friedländer, *op. cit.*

² Fernberger, *Amer. J. Psychol.*, 1914, 25, 538ff. (*op. cit.*).

³ Fernberger, 'Interdependence of Judgments within the Series for the Method of Constant Stimuli,' *J. Exper. Psychol.*, 1920, 3, 126-150.

5. Lest the argumental sauce obscure the meat, let me summarize. I recommend (1) that every judgment in the psychophysical experiment stand absolutely independently in its own right. One member of the series must be separated from the others in time, and by the instruction to the subject that he judge it without reference to any other member. Haphazard presentation should be the rule; at any rate serial presentations with partial knowledge should not be allowed since they connect the members and interfere with their individuality. I urge further (2) that the criteria of judgment be laid down explicitly and univocally in psychological terms. These psychological terms will be sensory in the class of experiments especially under consideration. Judgments of stimulus are often desirable, but they are not the final ideal since they are *ipso facto* equivocal.¹ The univocal character of the criteria must be tested by introspection and by the subjects' report. (3) The total *Aufgabe* under which the subject judges must be made definite in instructions, and must be more fully determined by means of repeated characterizations by the subject of his attitude and procedure. This latter check is important since much of the subject's instruction is apt to be a self-instruction. (4) The subject must be both instructed and trained to maintain a constant attitude throughout the experiment and to report lapses from this attitude. When he has learned the full meaning of this instruction, he will not give doubtful judgments nor ordinarily be doubtful, provided his task is made sufficiently easy for him by the means of the three foregoing rules. He will in like manner avoid other reflective judgments that violate the constant attitude. (5) He will probably tend to give immediate judgments, and he will be greatly helped if he is encouraged to report quickly. His times, however, will vary

¹ I have met psychologists who smile superiorly when I mention the 'stimulus-error' and even have something to say in reply about 'bigoted introspectionism,' so I know that I ought not, on purely diplomatic grounds, to bring the stimulus-error in; but unfortunately for diplomacy it belongs in. I hope some day to show that the stimulus-error is not a figment of an epistemologizing or a quibbling mind, but that it is a very real scientific devil. In the meantime let those whom numerical measures alone will impress, see Friedline, and Friedländer, *opp. cit.*

and there will occasionally be long delays [they seem like 'inhibitory jams'] without gross shift of attitude.

Undoubtedly research will bring more means of control to the fore, but the observance of these five rules alone will, I think, give Dr. Thomson data for which the thresholds will show the degree of constancy that he desires. If in the course of doing all this he thinks that we have made the thresholds 'more physiological' and less 'moral,' well and good. It is of this sort of stuff, nevertheless, that psychology is made.

THE NATURE OF THE PSYCHOMETRIC FUNCTIONS

In the article which has caused me to write this paper, Dr. Thomson supplies us with certain suggestions as to the nature of the psychometric functions—an exposition that employs the familiar device of the urn and balls. He gains comfort from this analogue, I take it, because within it he can show how the number of black balls necessary for a given category 'depends upon a conscious act of the subject, and can be varied, if he be so disposed, at his whim.' He gains support for his analogue (1) from the fact that it gives psychometric functions that are in accord with present statistical analysis¹ in that they are not normal curves (they are not the *phi*-function of *gamma* nor the normal bell) and (2) from the fact that the three psychometric functions are founded upon a single underlying error curve. I am quite ready to be convinced of (1), but (2) seems to me scarcely a reason, since there are other ways of founding a set of psychometric functions upon a single error function. I have suggested such a derivation elsewhere,² and I desire here to raise the question which of these two analogues better represents our notion of the psychophysical organism, and, furthermore, whether experiment and curve-fitting may not ultimately decide between the two and thus throw light upon the nature of sensitivity.

¹ Cf., Thomson, *op. cit.*, 304-307; 'The Criterion of Goodness of Fit of Psychophysical Curves,' *Biometrika*, 1919, 12, 226-229.

² Boring, 'A Chart of the Psychometric Function,' *Amer. J. Psychol.*, 1917, 28, 465-470.

Dr. Thomson's parable runs thus. A stimulus (or a stimulus-pair) is like an urn with black and white balls in it in a given proportion. The experimental trial is the drawing of a given (constant) number of balls from the urn. Something varies so that different numbers of black balls turn up in the successive drawings. The number of black balls drawn represents the impression; there is a series of impressions possible for a single stimulus all the way from no black balls to all black balls. (At least so I interpret Dr. Thomson.) The subject reports on this series in terms of (say) three categories. Dr. Thomson thinks that the subject decides, 'if he be so disposed, at his whim' what proportions of black balls drawn shall be reported by each of the categories, *e.g.*, whether 40 per cent. to 60 per cent. shall be reported as 'undecided' with the reports 'less' for under 40 per cent. and 'greater' for over 60 per cent. I urge that rigid criteria should replace the subject's whim. The same thing happens for the other stimuli, which are represented each by other urns in which the proportion of black balls is different. The series of urns, analogous to the series of stimuli, shows a continuously increasing (or decreasing) proportion of black balls, but the relation between the place of the urn in the series (stimulus-value) and the proportion of black balls need not be linear. Psychometric functions result that are similar to those actually obtained in practice, although they are not normal functions.

My own parable, adapted to the differential limen and the urn, separates the stimulus from the organism. A given stimulus supplies a fixed number of black and white balls, since a stimulus is ideally constant. In stimulating the organism its effect is facilitated or inhibited according to the chance disposition of the organism, that is to say, the organism is an urn from which an additional number of black and white balls is drawn to be added to the fixed number determined by the stimulus. The total resulting proportion of black balls fixes the position of the impression in the impressional series, and the report in terms of predetermined categories occurs as it does in Dr. Thomson's reasoning. The constant

tendency toward a given proportion of black balls is a function of the magnitude of the stimulus, and the variability about this constant tendency is a function of the organism (the urn). The proportion of black to white balls in the urn may be anything at all but is the same for all stimuli. The resulting ogive psychometric functions may, in a certain simple case, be normal, but there is no reason why they should be nor *a priori* presumption that they would be.

My first question is this. Is it not more reasonable to ascribe constancy to the stimulus and variability to the organism, and to assume a law of physiological variability that is fixed and independent of the magnitude of the stimulus? In other words, should not the organism be represented by a single urn with contents of fixed composition? I confess I am not clear as to what Dr. Thomson's urns symbolize nor as to precisely where in his scheme variability occurs. He has an urn of different composition for every stimulus. If variability resides in the organism, then the law of variability changes for every stimulus, and such an occurrence does not seem to me physiologically understandable.

To my second question I am not even prepared to suggest an answer. May it not be possible to determine empirically what urn-and-ball analogy best fits the facts, and thus analytically to learn something new of the laws of organic variability? A determination would be ever so much better than guesses—Dr. Thomson's or mine,—but for most psychologists curve-fitting is precarious work. If we had the proper data, would the mathematical solution be feasible? Perhaps Dr. Thomson will tell us.