Appetite 53, 434-437. http://dx.doi.org/10.1016/j.appet.2009.08.004
submitted manuscript with a few verbal corrections

Lines, dashed lines and "scale" ex-tricks. Objective measurements of appetite *versus* subjective tests of intake.

DAVID A. BOOTH

Food Quality and Nutritional Psychology Research Group, School of Psychology, College of Life and Environmental Sciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, U.K.

Abstract

Investigators of appetite for food have been tricked into the twin illusions that ratings of the disposition to eat are subjective and amounts eaten at meals are objective. The reality is the opposite. Making a mark on a continuous or broken line specified by two levels of what the rater uses as a single concept is the objective performance of a quantitative judgment. In contrast, the amount of a test meal that a person eats is a completely subjective outcome accumulated from many choices of another mouthful, each subject to several rapidly changing influences. Hence, rather than intake at test meals providing any validation for ratings of appetite, measurements of effects on the judged disposition to eat available food at each moment during a meal are needed to explain the amount consumed. This short paper is written in the hope of exorcising such self-deception from the research community and restoring systematic ratings of appetite to the uses for which they were introduced 35 years ago.

Key words: appetite ratings; food intake; test meals; objective verbal judgments; subjective amounts eaten.

Corresponding author
Email address: D.A.Booth@Bham.ac.UK (David Booth)
[Not for publication: Tel 0121 414 4938, Fax 4897]

How much or little do you want to eat or drink right now?

Ratings of appetite for foods and beverages have been widely misused over the last 30 years to perform various sorts of intellectual trickery. This brief paper aspires to expunge such deceptive practices from research into eating and drinking.[1]

The paper's main title adapts the phrase popularised by Mark Twain, "lies, damned lies and statistics." The present condemnation includes undefined scores from 100-mm lines, whether continuous or dashed, but extends to all bogus claims as to what is measured by responses on any layout.[2] There is no warrant in just the scores from any one rating for the common assertions that the rater is using any particular psychological scale or estimating the subjective magnitude of a sensation, a pleasure or a desire.

These errors have nothing to do with statistics. They are failures in scientific measurement. The issue is about validity. Until we know what we are measuring, questions about the reliability of our numbers make no sense.

Ratings to the rescue of research on intakes

Experimental analysis of human appetite for food (and, more rarely, for water) began by using a technique transferred from research on laboratory animals - weighing how much of the material(s) provided by the investigator disappeared into the eater's mouth during a session. This "intake" was presumed to be the visual analogue scale (*sic*) of the appetite for a genuine meal (or drink).

At the start, it was implicitly recognised that monitoring amounts that people consume is a dumb way to study the mechanisms of appetite. Choices of what to put in the mouth are driven by attitudes and expectations. Hence what is in mind has to be monitored somehow, or at least stopped from confounding whatever influences are being investigated. The total amount eaten reveals nothing about the changing influences over eating within, before or after the test period. The only aspect of intake observable within the meal is rate of eating and this yields little or no information about influences on eating as they operate mouthful by mouthful (Booth, Lee & McAleavey, 1976; cp.Yeomans, 2000).

Therefore, in a series of student projects reported in 1975-7 at widely attended meetings in Britain, France and the USA, we began to ask eaters what they could say about what was going on at the present moment that was relevant to taking a mouthful of food and also to make quantitative judgments about each observable feature of the situation at the moment of rating.

Our first step was to create as wide a variety as we could of wordings and formats, in order to find the rating that was most sensitive to disconfounded somatic, social and sensory sources of influence on appetite. Unfortunately others presumed that each wording that they chose measured what they thought it did, be it the desire to eat, a private bodily sensation associated with wanting to eat or ceasing to eat, a wish to eat food in a "large" amount (whatever that is), a thrill of pleasure from the taste of the food, or whatever. This fallacy of direct scaling is exposed whenever the responses to the different wordings are factor-analysed. Such scientifically basic reduction of data demonstrated from the start[4] that a single variable underlay all the ratings: each had a high loading on one component (except sometimes the rated acceptabilities of named foods in a common meal such as breakfast). That is, whatever words are used in a question relevant to the current disposition to eat, the answer expresses appetite for food (Booth, 1976).

Hence the objective analysis of influences on eating requires as the dependent variable only a numerically scorable answer to a question about current appetite for food. Mechanistically uninterpretable tests of amounts eaten are a waste of resources.

Inability of intake to measure hunger, thirst or satiety

Tests of intake are entirely subjective

Amount eaten at a meal is almost universally regarded as the "objective"

standard on which ratings of appetite for food must be "validated." The truth is the exact reverse of this. Conclusions from weights of foods consumed about appetite for food (or water), satiety, preference, acceptance, pleasure, wanting or anything else to do with eating (and drinking) need to be validated by measurements that use words quantitatively within test situations that are adequately designed and analysed.

Those who privilege intake testing over appetite rating fail to notice the utter subjectivity of the measure they rely on.

The status of weight of food as a physical quantity is transferred illegitimately to what it is claimed to measure - namely, factors controlling those weights. Even at the same material level, the chemical senses are stimulated by the concentrations of compounds in foods, not their weights, and gastric distension reflects volume not weight.

Worse, the concept of ad-lib intake, or feeding freely, is applied to the test meals as though it could mean in human beings (or indeed a species living in the wild) anything like what it does in an animal confined to a cage with continuous access to unlimited food. The instruction to eat what you want or as much as you like ("*ad libitum,*" *sic*) leaves it entirely to the eater to decide how much to eat. This completely destroys the pretension that the size of a "test meal" is an objective measure of anything behavioural.

What do the eaters want from the food? Is it to satisfy their wildest dreams, to save them buying supper, to support or undermine some hypothesis they believe the experimenter is testing or to create a pleasant sensation in the midriff? The most sensible aim for a test might be that the foods on offer are eaten as usual. Even if that instruction is given, it remains extraordinarily vague. What is the eater supposed to simulate – a meal on those foods or one on any foods? - at that time of day or when they "usually" eat (such food)? - alone or in company, at home or out, or what?

No attempt is made to provide the participant any sort of criterion by which to decide to stop eating, not even "what you would usually eat of such food at this time of day," let alone reference to an actual recent meal that the experimenter has tried to mimic, or to the specific influence(s) of interest, such as what is expected by the eater's primary other, what yields a satisfyingly full sensation or when all interest fades in the foods on offer. This is feeble science indeed compared to a properly structured answer to a precisely phrased question as to the likelihood that the eater would eat a mouthful of a specified food if given the opportunity right then in the exact circumstances of making that assessment (Booth, 2009b).

Meal size as a non-verbal rating

Those who think that weighing the foods makes intake at a meal into an objective measure have failed to recognise a central feature of human performance. The decision on how much to eat on an occasion is a quantitative judgment, just like any rating. If the size of the test meal intake is controlled at all, rather than the amount of food eaten being an arbitrary number, the judged quantity has to be by comparison with criterial levels of each of the operative influences. That is, there are some aspects of a standard meal implicit in memory at the moment(s) of decision. If nothing is extracted from the data on the tested situation about any such aspect, its level in the standard and the precision with which it is distinguished from the situation, then that test of intake may contribute nothing to knowledge about appetite.

For example, since visible volume dominates choices of amounts to eat (Jordan *et al*., 1969; Pudel, 1971), the physiological factor most likely to influence the size of a meal is the volume of food in the stomach. Comparison norms of gastric volume can be learnt from nutritional and social consequences, as in simulations by Booth, Toates and Platt (1976; Booth & Mather, 1978) that account for breakfast, lunch and supper. Whether in someone who is trying to eat less in order to lose weight or a non-dieter, the norm for amount in the stomach from a particular type of

meal is no more or less physiological or cognitive than any other learnt norm, such as the saltiness or sweetness of a particular food (Booth, Thompson & Shahedian, 1983, Conner & Booth, 1988) or the health implications of a calorie or fat label (Freeman *et al*., 1993). Such cognitive analysis of mechanisms controlling intake (Booth, 2008) has indeed now been extended to the amount in the stomach discriminated from a lunchtime norm for the test food (Kissileff *et al*., 2008)

A norm-based decision does not necessarily produce a meal of the usual size. For example, after we have been eating for some time, the comparison with a standard gastric volume may tell us that we are overfull to some degree, instead of comfortably satisfied. We may attribute that overeating to an emotional state, tempting food, indulgence in a treat, or reasons for giving up on a diet – long postulated mechanisms, although strangely little investigated. The meal could still be under the influence of discrimination from a gastric norm established while out of such a state but we may learn a new norm, opening up the risk of a bingeing habit. The gastric volume norm certainly can be shifted, for instance by the aversive after-effect of concentrated maltodextrin (Booth, 2009a), in other species as well as in us.

The invalidity of tests of intake

Amount eaten has other problems as a measure of appetite. These include reactivity – the operation of measuring having an effect on what is meant to be measured. There is also indeterminacy in the measure of amount, particularly for meals made up of several foods: weight can be a fair surrogate for volume because most foods have a density of about 1 $g.ml^{-1}$ but the usual reports in units of energy bury those data and presuppose that energy contents control meal sizes, a principle disproved long ago (Wooley, Wooley & Dunham, 1972).

Above all is the logical incoherence in the notion that a bout of eating provides a valid measure of the state of appetite. The influences on appetite not only vary within the test meal but also between meals. Yet intake is measured without regard to the levels of those influences at the time of testing.

So much for the notion that intake is something that appetite ratings need validating on. May we please dispense with it? The only relevant validity and hence reliability is that of the effects on rating the momentary strength of appetite of one or more influences that are disconfounded from each other and from other factors.

Objective measurement of appetite using ratings

States of hunger, thirst and their satieties

Users of ratings need to get to grips with fact that they are asking the rater to perform an objective task of judging the quantity of something that actually exists in the publicly observable world. The rater's achievement cannot be determined merely from the scribbles on a piece of paper (or their digital analogues on a screen).

The fact that the rater's response by itself measures nothing is not changed by an investigator's belief that it estimates a subjective magnitude, neural processing somewhere in the brain or a social construction. Furthermore, none of these claims can be right because there is logically no way of deciding among them from data on one wording in response to one type of stimulus.

What the scores extracted from the responses actually succeed in quantitating can be ascertained only by measuring the information coming in from the observable environment that is transformed within the rater's mind into information that is imposed on that material and social environment. That transformation of input into output is the rater's dynamic mental state at the moment that the quantitative judgment is being expressed on a verbally anchored sequence of positions or numbers.

An integrative act, such as rating appetite or taking a mouthful, results from conscious and unconscious interactions among stimuli, memories and options that can be characterised by appropriate analyses of data from a correctly designed test

situation (cp. Booth, 1989, Table 2, and Booth & Freeman, 1993). The minimum requirement is at least two disconfounded levels of each of two or more material and/or symbolic features of the eating situation. The characterisation of the momentary mental state is made more specific by the addition of at least one rating that is 'analytical' – that is, shown from the data to respond only to one influence on appetite (cp. Booth *et al*., 2003). That also makes it possible to distinguish conscious sensation from implicit perception, mere verbalisation, correct description or active emotion and attitude (Booth & Freeman, 1993; Galea *et al*., 2008).

Objective ratings of appetite used subjectively. Scores for hunger, fullness, pleasantness of foods, etc., have been called "subjective appetite" (and satiety) in a contrast with intake of food regarded as an "objective" measurement of appetite and its sating. If the distinction needs to be drawn between symbolic and material data relevant to eating, the correct term is "rated appetite." (It is a solecism to refer to "subjective ratings" of appetite, satiety or whatever: if verbal or pictorial data are wrongly regarded as subjective, then there is no sense in saying that appetite was assessed both by ratings and subjectively.)

The only thing "subjective" about ratings of appetite is many investigators' uses of the data. Quantitative judgments of strength of tendency to eat are a crucial part of the basic apparatus for investigating the causal processes during an episode of eating and drinking. Any number extracted from a graded way of answering a question how hungry or full you are right now, or how pleasant it is to eat a mouthful of a food item, is *prima facie* a valid measure of what the rater regards as appetite for food. However, mere recordings of how strong people say appetite for food is at a given moment do not contribute to knowledge, any more than do mere weights of food eaten. Scientific issues about dispositions to eat arise only with investigation also of the processes influencing what a person eats and drinks, and in what amounts.

Research begins when one raw score (or a combination score representing differently worded ratings) is related to quantitative or categorical variations in a monitored feature of the social situation (e.g. ambience, another person) or material context (e.g. a food, contents the stomach, labelled contents of a pack). Then the scores for appetite could be measuring some influence on eating which is at least partly characterised.

Rating appetite of the moment. Silverstone and Stunkard (1968) scored strength of hunger simply as a position on a line from 'not at all hungry' to 'extremely hungry,' in order to measure the effect of an appetite-suppressant drug. An inventory of questions is useless unless subscaled, i.e. different sets of items have been shown to measure different variables with the context of each experiment.

It is a mistake to ask direct questions about the strength of a particular influence on eating – for example, "how pleasant is the sweetness of this yoghurt?" The integrative rating of overall appetite, "how pleasant is this yoghurt?" needs to be on a separate subscale from the analytical rating, "how sweet is this yoghurt?"

It is also a mistake not to anchor the answer to any question on observables – materials, labels or actions. An integrative rating should be anchored on definite acts, e.g. "I'd always choose it" and "I'd never choose it" (or worse). An analytical rating can be anchored on such actions or on a label for a familiar version of the material, e.g. "as sweet as my usual yoghurt [of this flavour]."

Then the influence of whatever is referred to by the rater as "sweetness" on the disposition to eat that yoghurt at the time of rating is measured by how good the pleasantness ratings are at discriminating between the sweetness ratings (Booth & Freeman, 1993). Also, as well known but little used in industry (Booth & Conner, 2009), the personal ideal of sweetness in a food or drink can be calculated from closeness to "always choose" of sweetnesses relative to usual.

If the concentration of a sugar or intense sweetener in the yoghurt is varied

among samples, then the personally ideal physical level can be calculated (Booth *et al*., 1983; Conner & Booth, 1998). What is not generally realised among psychophysicists is that such data can also address the otherwise unresolvable controversies whether sensory influences on appetite involve sensations, unconscious neural processing, use of the concept 'sweet,' just wanting to eat, or indeed successful description of the sweetener as sweet (Booth & Freeman, 1993). The same can be done for feeling "full" when amount in the stomach is varied independently of the amount visible or swallowed (report in preparation, cf. Kissileff *et al*., 2008).

Long-term appetite

Appetite as a symptom. A clinician may want to consider a patient's 'average' appetite. Parents (and those who do research into children's eating) may refer to a child as having more or less of an appetite. Farmers (and farm animal scientists) use the word "appetite" to refer to an animal's usual daily intake. These usages are legitimate if they can be shown to explain other data. It would be a mistake for a clinician or a parent to seek a number for total intake, since the best dietary assessment cannot do that for the individual. An impression of poor appetite should being regarded as a symptom – a measure of status rather than of state: then the rating of appetite needs to be validated on the disease or disorder of which it is supposed to be diagnostic.

Traits of appetite. Sets of questions about aspects of appetite may have a role if they can be shown by orthogonal factor analysis to tap distinct variations among individuals in habitual aspects of appetite in everyday life. Traits can be hard to separate from states: insistence on answers about some future time or that cover some hours or days into the past does not stop the answers being influenced by current state (as in any so-called "trait measure").

Stunkard and Messick (1985) collated a large number of questions about the desire to eat and constructed a generic multi-item scale that they dubbed 'Hunger.' Their questionnaire gave two other multi-item scales, that they called 'Restraint' and 'Disinhibition'. All three scales were meant to predict overeating. However, it is far from clear that distinct mechanisms are tapped by subscales for emotional eating, susceptibility to tempting foods (disinhibition) and efforts to cut back on calories (restraint).

Hunger and obesity. A popular use of the notion of a general level of appetite for food in a person is in relation to obesity. Being seen to eat a lot or to eat fast may not however explain or presage obesity, unless total intake is increased over the long term after such eating. Hence, appetite that fattens can never be measured by tests of effects on meal size, however often the effect holds up. Similarly, hunger or satiety at any one moment has zero implications for obesity.

Other erroneous sorts of generic concept (Booth, 1990) include palatability (appetite for a food that remains constant across contexts) and satiating power of a nutrient, without regard to delay after consumption, amount consumed, proportion to other nutrients and so on. Both these notions have been invoked to explain obesity. This is as plausible as using current mood to measure chronic anxiety or depression.

The only way to measure long-term fattening hunger or slimming satiety (Booth, 1996) is to track individuals' frequencies of local eating customs and reported body weights and to measure from the results the effect on weight of change in frequency of each custom (Blair *et al*., 1989; Booth *et al*., 2004; Lewis, Blair & Booth, 1992). Once again, the only valid measure of the variations in appetite that result in energy intake comes from eaters' words, not any weights of foods.

References

Blair, A.J., Booth, D.A., Lewis, V.J., & Wainwright, C.J. (1989). The relative success of official and informal weight reduction techniques. *Psychology and Health* 3, 195-206.

Booth, D.A. (1976). Approaches to feeding control. In T. Silverstone (Ed.), *Appetite and food intake*. *Dahlem Konferenzen Life Sciences Report 2*. pp. 417-478. West Berlin: Abakon Verlagsgesellschaft.

Booth, D.A. (1990). How not to think about immediate dietary and postingestional influences on appetites and satieties. *Appetite* 14, 171-179.

Booth, D.A. (1996). Measurement of hunger and satiety for long-term weight control. In A. Angel *et al*. (Eds.), *Progress in obesity research: 7*, pp. 355-358. London: John Libbey.

Booth, D.A. (2008). Physiological regulation through learnt control of appetites by contingencies among signals from external and internal environments. *Appetite* 51, 433-441.

Booth, D.A. (2009a). Learnt reduction in the size of a meal. Measurement of the sensory-gastric inhibition from conditioned satiety. *Appetite* 52, 745-749.

Booth, D.A. (2009b). The basics of quantitative judgment. How to rate the strength of appetite for food and its sating. *Appetite* 53, xxx-xxx *[Setter: please take pagination from accompanying Short Communication.]*

Booth, D.A., & Freeman, R.P.J. (1993). Discriminative measurement of feature integration in object recognition. *Acta Psychologica* 84, 1-16.

Booth, D.A., & Conner, M.T. (2009). Salt in bread. [Letter to the Editor] *Journal of Food Science* 74(3), vii-viii.

Booth, D.A. & Mather, P. (1978). Prototype model of human feeding, growth and obesity. In D.A. Booth (Ed.), *Hunger models*, pp. 279-322. London: Academic Press.

Booth, D.A., Blair, A.J., Lewis, V.J., & Baek, S.H. (2004). Patterns of eating and movement that best maintain reduction in overweight. *Appetite* 43, 277-283.

Booth, D.A., Lee, M., & McAleavey, C. (1976). Acquired sensory control of satiation in man. *British Journal of Psychology* 67, 137-147.

Booth, D.A., Mobini, S., Earl, T., & Wainwright, C.J. (2003). Market-optimum instrumental values from individual consumers' discriminations of standard sensory quality of the texture of short-dough biscuits. *Journal of Food Quality* 26(5), 425-439.

Booth, D.A., Thompson, A.L. & Shahedian, B. (1983). A robust, brief measure of an individual's most preferred level of salt in an ordinary foodstuff. *Appetite* 4, 301-312.

Booth, D.A., Toates, F.M., & Platt, S.V. (1976). Control system for hunger and its implications in animals and man. In D. Novin, W. Wyrwicka & G.A. Bray (Eds.), *Hunger: basic mechanisms and clinical implications*, pp. 127-142. New York: Raven Press.

Conner, M.T., & Booth, D.A. (1988). Preferred sweetness of a lime drink and preference for sweet over non-sweet foods, related to sex and reported age and body weight. *Appetite* 10, 25-35.

Freeman, R.P.J., Richardson, N.J., Kendal-Reed, M.S., & Booth, D.A. (1993). Bases of a cognitive technology for food quality. *British Food Journal* 95 (9), 37-44.

Galea, J., Chechlacz, M., Booth, D.A., Higgs, S., Birbaumer, N., & Nouwen, A. (2008). Perceptual and affective processing in appetite for foods. *Appetite* 51, 367.

Jordan, H.A., Wieland, W.F., Zebley, S.P., Stellar, E., & Stunkard, A.J. (1969). Direct measurement of food intake. A method for objective study of eating behavior. *Psychosomatic Medicine* 28, 836-xxx

Kissileff, H.R., Booth, D.A., Thornton, J.C., Pi-Sunyer, F.X., Pierson, R.N., & Lee, J. (2008). Human food intake is discriminatively sensitive to gastric signaling. *Appetite* 51, 759.

Lewis, V.J., Blair, A.J., & Booth, D.A. (1992). Outcome of group therapy for body-image emotionality and weight-control self-efficacy. *Behavioural Psychotherapy* 20, 155-166.

Pudel, V. (1971). Food dispenser: method for study on spontaneous appetite behavior. *Zeitschrift für Ernährungswissenschaft* 10, 382-xxx.

Silverstone, J.T., & Stunkard, A.J. (1968). Anorectic effect of dexamphetamine sulphate. *British Journal of Pharmacology* 33, 513-xxx.

Stunkard, A.J., & Messick, S. (1985). The 3-factor eating questionnaire to measure dietary restraint, disinhibition and hunger. *Journal of Psychomatic Research* 29, 71-83.

Wooley, O.W., Wooley, S.C., & Dunham, R.B. (1972). Can calories be perceived and do they affect hunger in obese and nonobese humans? *Journal of Comparative and Physiological Psychology* 80**,** 250-xxx.

Yeomans, M.R. (2000). Rating changes over the course of meals: what do they tell us about motivation to eat? *Neuroscience and Biobehavioral Reviews* 24, 249-259.

Footnotes

1. The papers criticised are not cited here, in order to avoid giving more publicity to their advocacy of errors. Sad to say, these misconceptions have become so pervasive that sometimes they have survived peer reviewing even in this Journal. It is to be hoped that this will now not happen again. Also other journals might take note, especially in medical and industrial fields.

2. Scalextrics (formerly Scalectrics in Britain) is a brand of small toys, such as motorcars and motorcycles. Boys and grown men can get very excited playing with these toy vehicles but they remain fully aware, it is to be hoped, that their games bear no relation to real driving and racing.

3. Around that time, the orally intolerable blood glucose tolerance test using concentrated glucose was being switched to maltodextrin. This starch product has chains of glucose short enough to be soluble but long enough to avoid hypertoncity and so should be used as the carbohydrate instead of sugars (Booth *et al*., 1976).

4. In 1977, Peter Mather and I factor-analysed many days of a few eaters' ratings of appetite in a dozen ways (cp. Booth *et al*., 1982) in a nutrition trial run by Mike Tame at Unilever Colworth. After rotation, one factor always accounted for most of the variance. All the general ratings of appetite loaded heavily on it, plus usually the food-specific ratings for a frequent menu such as breakfast. The paper was submitted to *Appetite* for its first Volume in 1980. Unfortunately for the field thereafter, the reviewers were unshiftably stuck (as many still are) in the psychometric conventions for scales in clinical questionnaires, requiring data from questioning large numbers of people just once, although our use of factor analysis simply to reduce data is recognised nowadays in statistical packages.