R05-02

Appetite 44 (2005) 1–13

Research Review

# Development and testing of a labeled magnitude scale of perceived satiety

Armand V. Cardello[a,*], Howard G. Schutz[b], Larry L. Lesher[c], Ellen Merrill[c]

[a]*Science and Technology Directorate, US Army Natick Soldier Center, Natick, MA 01760-5020, USA*
[b]*Consumer Sciences Department, University of California, Davis, CA 95616, USA*
[c]*GEO-Centers, Inc. Natick, MA 01760, USA*

Received 29 July 2003; revised 30 April 2004; accepted 8 May 2004

ELSEVIER

2004

ELSEVIER

Appetite 44 (2005) 1–13

Appetite

www.elsevier.com/locate/appet

Research Review

# Development and testing of a labeled magnitude scale of perceived satiety

Armand V. Cardello[a,*], Howard G. Schutz[b], Larry L. Lesher[c], Ellen Merrill[c]

[a]*Science and Technology Directorate, US Army Natick Soldier Center, Natick, MA 01760-5020, USA*
[b]*Consumer Sciences Department, University of California, Davis, CA 95616, USA*
[c]*GEO-Centers, Inc. Natick, MA 01760, USA*

Received 29 July 2003; revised 30 April 2004; accepted 8 May 2004

**Abstract**

Satiety ratings are often made using VAS or simple category scales. In order to establish a simple, more quantitative technique to index perceived hunger and/or fullness, research was undertaken to develop and test a labeled magnitude scale of satiety.

Thirty-seven subjects rated the semantic meaning of 47 phrases describing different levels of hunger/fullness using magnitude estimation. Eleven phrases were then selected using criteria of response consistency, symmetry, bipolarity, and inclusion of the end-point anchors of 'greatest imaginable hunger (fullness)'. These phrases were placed along a vertical line scale at positions corresponding to their geometric mean magnitude estimates to create a labeled magnitude scale of satiety.

This Satiety Labeled Intensity Magnitude (SLIM) scale was compared to VAS scales for sensitivity and reliability in two studies. In one image-based study, ANOVA with post-hoc tests showed the SLIM scale to have greater sensitivity and to have an average reliability coefficient of 0.90. In a second study using three different foods replicated on two consecutive weeks, reliability was found to be highest for the SLIM scale. It is concluded that the SLIM scale is a sensitive, reliable, and easy-to-use scale for measuring perceived satiety that has several advantages over other, more commonly used satiety scales.


*Keywords:* Satiety; Hunger; Fullness; Scaling; Labeled magnitude scales

## Introduction

US Army soldiers operating in combat training exercises underconsume their rations, resulting in a 10% weight loss after several weeks in the field (Hirsch, 1995; Hirsch & Kramer, 1993; Meiselman, Hirsch, & Popper, 1988; Meiselman, 1995). From an historical perspective, the military has focused its efforts to increase consumption on improving the sensory quality and acceptability of the rations. However, the reasons for underconsumption in the field are more varied and complex. In addition to any contribution from the sensory quality of the food, reasons commonly cited include the time and effort required to properly prepare and consume the rations, the poor situational and environmental conditions in which they must be consumed, negative attitudinal factors that foster lower consumption, poor hydration status, and reduced food variety. These reasons have led military investigators to conclude that consumption in a natural eating environment is greatly dependent upon 'the situational variables which make it more or less convenient for us to eat and which signal meal time' (Meiselman et al., 1988).

The realization that underconsumption of rations may be less dependent on the sensory quality of rations than on the conditions under which they are eaten has led the US Army Natick Soldier Center to explore other strategies for improving field consumption. One strategy currently being considered is to design rations to have higher caloric density while maintaining or reducing their satiating capacity, so that larger caloric quantities must be consumed in order to achieve the same level of perceived satiety. In order to achieve this latter goal, a better understanding is needed of the differential capacity of foods to induce feelings of fullness and of the physicochemical, sensory, and macronutrient-related factors that contribute to perceived satiety.

* Corresponding author.
[redacted]

0195-6663/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.appet.2004.05.007

In a recent series of papers, Holt, Miller, Petoez, and Farmakalidis (1995) and Holt, Miller, and Stitt (2001) published satiety indices for different foods. These indices were based on studies in which subjects ate isocaloric portions of different foods and their satiety was tracked over time using a subjective rating scale. Upon examining Holt et al.'s methods, it was apparent that the accuracy and utility of their indices were greatly dependent on the subjective measure of satiety that was used. In their studies, Holt et al. used a horizontal, 100 mm visual analogue scale that was verbally anchored at the end-points and at five equally spaced intervals between the end-points, creating a seven-pt scale. From left to right, the seven verbal labels were 'extremely hungry', 'hungry', 'semi-hungry', 'no particular feeling', 'semi-satisfied', 'satisfied', and 'extremely full'. Unfortunately, from a psychophysical standpoint, this scale suffers from a number of problems. First, the scale is a 'mixed scale', that is, it utilizes a variety of different terms, e.g. hungry, satisfied, and 'feeling' that do not necessarily represent a single measurement dimension. Secondly, the term "satisfied" introduces a hedonic component into what is meant to be an intensity (hunger/fullness) dimension. Thirdly, the psychophysical advantage of a visual analogue scale, i.e. that the scale is continuous and unperturbed by non-equivalence of scale-point intervals, is violated by the placement of verbal labels representing unknown magnitudes of satiety at equally spaced intervals along the scale. These labels and the mixed nature of them transform the scale into a hybrid scale, for which the assumption of linearity of response is violated and for which even the assumption of providing ordered metric data may be suspect.

In a recent study (Merrill, Kramer, Cardello, & Schutz, 2002), the Holt et al. (1995) scale was compared to both uni-directional and bi-directional visual analogues scales (VAS) of hunger, 'fullness', and 'amount could eat' in order to assess the perceived satiety of foods. The results of this study showed that the amount could eat scale and the unipolar fullness scale were generally less reliable and less sensitive to differences in satiety produced by the different foods than the other scales tested. Although the Holt scale was not as sensitive as a simple bi-directional 'hunger-fullness' VAS scale, its average reliability coefficient was the highest among the scales. This latter finding is consistent with its numerous, anchored verbal labels, which would aid memory and reduce the variability inherent in repeatedly placing a mark along an otherwise, unstructured line scale.

As a result of our inability to identify an optimal scale for assessing satiety from among existing scales, we decided to undertake the development of a psychophysically improved scale of satiety that could be used in subsequent studies to measure the relative capacity of different ration components to induce satiety.

Since the early work of Silverstone and Stunkard (1968), numerous investigators have used VAS scales to index the sensory and perceptual dimensions of human satiety, hunger, fullness, desire to eat, etc. (e.g. Blundell & Burley, 1987; Burley, Leeds, & Blundell, 1987; Green & Blundell, 1996; Hill, Magson, & Blundell, 1984; Kisseleff, Gruss, Thornton, & Jordan, 1984; Robinson, Folstein, & McHugh, 1978; Rolls et al., 1994, 1999). The use of VAS scales has been supported by data showing that such measures are predictive of subsequent food intake (Barkeling, Rossner, & Sjoberg, 1995; Flint, Raben, Blundell, & Astrup, 2000; Hulshof, de Graff, & Westrate, 1993; Porrini, Croveti, Tesolin, & Silva, 1995; Robinson, McHugh, & Folstein, 1975; Silverstone, 1966). VAS scales also avoid measurement problems commonly associated with simple category scales of satiety that have been used in the past, e.g. Wooley and Wooley (1973). The latter scales suffer from two significant problems. The first is that the labeled points on the scale rarely define equal intervals. The reason for this is that the verbal phrases are rarely examined or selected on the basis of quantitative data concerning the psychological magnitudes that they express. This is true for even well documented and commonly used category scales, like the nine-pt hedonic scale (see Moskowitz, 1977; Schutz and Cardello, 2001). This non-equivalence of scale intervals reduces the mathematical level of the data obtained from interval data to what is, at best, ordered metric data. In spite of this, most investigators continue to treat data obtained using category scales in a mathematically inappropriate manner by applying parametric, rather than non-parametric, statistics to the data.

The second problem with category scales is a 'central tendency' or 'regression' effect that results in under-use of the end categories (Stevens & Galanter, 1957). Subjects avoid the end categories, because once they assign a sensation to one of these categories, subsequent sensations that are even more intense (in either direction) cannot be accommodated. This avoidance of the end categories effectively reduces the number of scale points and limits the ability of the scale to discriminate among different intensity levels of more extreme sensations.

VAS scales avoid the above problems, but the scale type often becomes indeterminate because the subjects are rarely instructed on how to make their judgments on such scales. Such scales are not ratio level scales, because they have fixed end-points. In addition, unless instructions are offered to subjects that their responses should be made in a ratio manner to one another, these scales do not produce ratio data. Methodological issues related to the validity, reliability and sensitivity of VAS and other scales of satiety have been reviewed and discussed previously (Booth, 1981; deGraaf, 1993; Hill, Rogers, & Blundell, 1995; Kisseleff, 1981; Stubbs et al., 2000; Stubbs, Johnstone, O'Reilly, & Poppitt, 1998). Several of these discussions were precipitated by the introduction of a form of ratio scaling, i.e. cross-modality matching, to the assessment of the magnitude of hunger/fullness (Teghtsoonian, Becker, & Edelman, 1981). These latter researchers suggested a method for scaling fullness that required subjects to stretch a tape measure to a desired length, such that the drawn length was proportional to the fullness that they were experiencing at the time of their judgment. Based on previous psychophysical evidence, such a scale should produce ratio level data (Stevens, 1971) and allow statements to be made about the ratios of fullness experienced by subjects. Such statements cannot be made with ratings made on a VAS scale.

Ratio scale methods were first introduced by S.S. Stevens (S.S. Stevens, 1957; Stevens, 1962; Stevens & Galanter, 1957). He and his co-workers developed a number of ratio scaling techniques, including cross-modality matching and magnitude estimation. In magnitude estimation, sensations are scaled by having subjects assign numbers to each sensation in such a way that the ratio between assigned numbers is the same as the ratio between the sensations they represent. This technique, coupled with the use of a true zero point for null sensations establishes a ratio scale for the data, enabling the researcher to conclude that one sensation is twice (one-third, three-fifths, etc.) as intense as another sensation. Although this method provides the highest possible level of measurement, it, like other common ratio scaling techniques, is cumbersome to use and requires detailed instructions for untrained consumers. In addition, since all judgments are made relative to one another, there is no provision for anchoring judgments of individual subjects to a common, absolute ruler. Although some of these practical problems may be addressed by utilizing ratio instructions with a linear graphic or VAS scale (Lawless, 1977), the requirement to accommodate theoretically infinite magnitudes and the instructional difficulties associated with having subjects judge ratios of intensity, still pose large practical obstacles. For these reasons, ratio scaling techniques have not been widely adopted outside of the mainstream of psychophysics.

In an effort to combine the practical simplicity of category scales with the quantitative advantages of magnitude estimation, Green, Shaffer, and Gilmore (1993) developed a 'labeled magnitude' scale for rating the intensity of oral sensations. This scale, modeled after the 'category-ratio' scale developed for the measurement of perceived exertion by Borg (1982), placed verbal labels of expressed intensity along a linear graphic scale at specific locations that reflect the numerical ratios among their perceived intensities, as derived from magnitude estimation procedures. A critical characteristic of such 'labeled magnitude' scales is the presence of a phrase involving the 'maximal', 'strongest imaginable' or 'strongest possible' sensation. The use of these phrases is assumed to serve as a fixed end-point of sensation that aligns judgments of different subjects to a common sensory 'ruler' and avoids the practical issue of how to accommodate extremely large magnitudes on a line of fixed length. With minimal instructions, subjects can quickly look at the verbal labels and corresponding numbers and place a slash mark (/) through the line to indicate the perceived strength of their sensation. The resulting data have been shown to produce psychophysical functions identical to magnitude estimation (Green et al., 1993). More recent studies of this scale (Green et al., 1996) have shown it to be most effective for application to chemosensory stimuli that include extremely intense (painful) sensations.

During the past several years, Bartoshuk and co-workers (Bartoshuk, 2000; Bartoshuk, Duffy, Fast, Green, & Prutkin, 2001; Duffy, Fast, Cohen, Chodos, & Bartoshuk, 1999; Prutkin et al., 2000) have argued forcefully for the use of labeled magnitude scales in all areas of sensory measurement, because of the ability that they offer to better compare sensations between individuals whose 'perceptual worlds' may be different due to greater or lesser sensitivity. Similarly, our laboratory has begun the development of labeled magnitude scales for affective dimensions, such as liking/disliking (Schutz & Cardello, 2001) and comfort (Cardello, Winterhalter, & Schutz, 2003) in order to improve the mathematical level of the obtained data.

The present report details the development of a labeled magnitude scale for assessing perceived satiety. The main objectives were to develop a scale that (1) is easy-to-use, (2) has reliability and sensitivity equal to or better than current category and VAS scales, and (3) provides ratio level data. Such a scale should enable greater discrimination of satiety sensations, especially at higher levels of hunger or fullness, and will enable ratio statements to be made about differences in the intensity of satiety sensations, e.g. 'twice as hungry', 'one-third as full', etc.

## Experiment 1

The purpose of the first experiment was to quantify the semantic meaning of common English phrases used to describe hunger and fullness, using modulus-free magnitude estimation.

### *Method*

#### *Subjects*

Fifty-eight male and female employees of the US Army Natick Soldier Systems Center were randomly selected from a larger pool of 350 individuals who had volunteered to participate in routine sensory tests of food. All panelists were familiar with food testing procedures. All had prior experience in using category scales to rate food acceptance. A small percentage ($<10\%$) had prior familiarity with the use of magnitude estimation.

#### *Stimuli*

A list of 47 phrases that can be used to describe different levels of hunger and fullness was developed from the satiety literature, English dictionaries, and from the general psychophysical scaling literature. These phrases are shown in Table 1. The phrases 'greatest imaginable' and 'greatest possible' hunger/fullness were included to define scale

Table 1
Phrases used in magnitude estimation study

| | |
|---|---|
| Greatest imaginable fullness | Neutral |
| Least imaginable hunger | Neither hungry nor full |
| Bursting | No particular feeling |
| Stuffed | Neither sated or unsated |
| Extremely full | No appetite |
| Very full | Semi-hungry |
| Gorged | Unsated |
| Surfeited | Slightly unsatisfied |
| Moderately full | Slightly hungry |
| Extremely sated | Moderately unsatisfied |
| Very sated | Hungry |
| Extremely satisfied | Empty |
| Very satisfied | Moderately hungry |
| Extremely content | Very hungry |
| Satisfied | Very unsatisfied |
| Slightly full | Extremely unsatisfied |
| Very content | Extreme appetite |
| Moderately sated | Extremely hungry |
| Moderately satisfied | Voracious |
| Moderately content | Ravenous |
| Slightly satisfied | Least imaginable fullness |
| Slightly sated | Greatest imaginable Hunger |
| Slightly content | |
| Semi-satisfied | |
| Neither famished nor gorged | |

values commensurate to common fixed end-points of hunger and fullness. Each phrase was printed on a separate sheet of paper (8×22 cm) along with space for making written magnitude estimates of their semantic meaning. The sheets containing each phrase were randomly combined into individual booklets for administration.

*Procedure*

All testing was conducted in individual testing booths. Panelists were provided a written set of instructions about the procedure to be used in scaling the meaning of each phrase (Appendix A). The procedure used was modulus-free magnitude estimation, in which subjects assessed whether the semantic meaning of the first phrase indicated some degree of fullness (+), hunger (−) or neither (0) and then assigned a number (free range) to indicate the magnitude of the hunger or fullness expressed by that phrase. The semantic meaning of all subsequent word phrases were then judged relative to the magnitude estimate assigned to this first phrase. In addition to written instructions, the details of the procedure were explained by the investigator(s) to each panelist, and all panelists were allowed to ask questions before starting. Each subject rated all 47 phrases for the perceived magnitude of hunger or fullness (or neither) expressed by the phrase.

*Results and discussion*

Since magnitude estimates have been shown to be log-normally distributed (J.C. Stevens, 1957) and since visual examination of the obtained data distributions showed deviations from normality typical of magnitude estimates, especially at the tails, the data were analyzed by equalizing the magnitude estimates across subjects (Lane, Catania, & Stevens, 1961) and then calculating the geometric means of the normalized magnitude estimates across subjects for each phrase. Positive and negative phrases were normalized separately, with absolute values used for the negative phrases. Several phrases were excluded from further consideration prior to normalization, based on the fact that they received positive, negative and zero ratings, indicating that they were highly ambiguous phrases. These phrases included 'least imaginable hunger', 'surfeited', 'unsated', and 'least imaginable fullness'.

Table 2 shows the geometric mean magnitude estimates for the phrases representing non-zero degrees of hunger or fullness. Examination of the data reveals the ratings to have general construct validity, because the rank order of geometric mean magnitude estimates corresponds to the generally understood and accepted semantic meaning of the phrases. Also, in keeping with previous findings concerning

Table 2
Geometric mean magnitude estimates for non-neutral phrases

| Phrases | Geometric mean magnitude estimates |
|---|---|
| Greatest imaginable fullness | 115.2 |
| Bursting | 107.1 |
| Extremely full | 91.5 |
| Stuffed | 91.3 |
| Very full | 85.6 |
| Gorged | 80.4 |
| Moderately full | 53.9 |
| Very sated | 50.0 |
| Extremely satisfied | 47.4 |
| Extremely sated | 45.9 |
| Very satisfied | 45.6 |
| Extremely content | 42.0 |
| Satisfied | 37.0 |
| Slightly full | 36.7 |
| Very content | 35.2 |
| Moderately sated | 35.1 |
| Moderately satisfied | 31.2 |
| Moderately content | 26.6 |
| Slightly satisfied | 19.7 |
| Slightly sated | 19.7 |
| Slightly content | 17.4 |
| Slightly unsatisfied | −21.1 |
| Slightly hungry | −23.8 |
| Moderately unsatisfied | −27.7 |
| Semi-hungry | −28.3 |
| Very unsatisfied | −36.8 |
| Empty | −42.7 |
| Extremely unsatisfied | −43.5 |
| Hungry | −47.8 |
| Moderately hungry | −48.9 |
| Voracious | −53.0 |
| Very hungry | −72.0 |
| Extreme appetitie | −72.1 |
| Ravenous | −81.8 |
| Extremely hungry | −86.3 |
| Greatest imaginable hunger | −107.1 |

the non-equivalence of intervals commonly assumed on labeled category scales, Table 2 demonstrates that among phrases using the terms 'hungry' or 'full', seemingly adjacent intensity modifiers do not constitute intervals of hunger or fullness that are perceptually equivalent. For example, while the interval between the phrases 'slightly full' and 'moderately full' is 17 units, the interval between the phrases moderately full and 'extremely full' is 25 units. In terms of ratios, 'extremely full' is 1.5 times greater than 'moderately full', while 'moderately full' is 2.1 times greater than 'slightly full'. Similar examples of non-equivalence of intervals can be seen for other pairs of adjoining phrases.

The present data show a reasonable degree of symmetry in the perceived magnitudes defining the end-points of hunger and fullness, as evidenced by the mean magnitude estimates of 115.2 for 'greatest imaginable fullness' and −107.1 for 'greatest imaginable hunger' (a 7% difference). However, the growth of the perceived magnitude of hunger or fullness for corresponding intensity modifiers, such as 'slightly', is not as symmetrical, i.e. slightly full = 36.7 but slightly hungry = −23.8 (a 35% difference).

In order to construct a labeled magnitude scale of perceived satiety from the data in Table 2, it was necessary to choose a set of phrases that spanned the full range of hunger and/or fullness values. The choice of using a bi-directional (hunger-fullness) scale, rather than two uni-directional scales (hunger only, fullness only) is always a consideration in satiety research. While the data generated in this experiment could be used to construct of any one of a variety of different labeled magnitude scales (both uni-directional and bi-directional), we elected to develop a bi-directional scale. Our reasons for doing so were (1) that the primary purpose for developing an improved satiety scale was to quantify the total perceived satiety produced by a food/meal, (2) that this purpose did not require dissociating subtle short-term changes in hunger and fullness, for which the use of both a hunger and a fullness scale may be beneficial (e.g. Herman, Ostovich, & Polivy, 1999; Yeomans, 2000), and (3) that our previous research had shown bi-directional scales to be generally more reliable and sensitive to differences in the total satiety value of different foods (Merrill et al., 2002).

Based on previous research on the development of related scales (Schutz & Cardello, 2001; Cardello et al., 2003), it was deemed desirable to (1) have a balanced number of phrases representing intensive levels of hunger or fullness, (2) ensure consistency in the word adjectives used to describe these levels of hunger and fullness, and (3) to use a common set of adjectives to describe the levels of hunger/fullness. With these criteria in mind the 10 symmetrical phrases of slightly hungry (full), moderately hungry (full), very hungry (full), extremely hungry (full), and greatest imaginable hunger (fullness) were chosen to construct the scale. In addition, a 'neutral' phrase was chosen to anchor the scale at the zero point. Since the phrases 'hunger' and 'fullness' had been chosen to describe the opposing dimensions of the scale, the neutral phrase 'neither hungry nor full' was chosen as the zero-point anchor.

Table 3
Point locations on a −100 to +100 scale for each of the verbal labels of the SLIM scale

| | |
|---|---|
| Greatest imaginable fullness | 100.0 |
| Extremely full | 79.4 |
| Very full | 74.3 |
| Moderately full | 46.7 |
| Slightly full | 31.9 |
| Neither hungry nor full | 0.0 |
| Slightly hungry | −18.6 |
| Moderately hungry | −38.2 |
| Very hungry | −56.2 |
| Extremely hungry | −67.4 |
| Greatest imaginable hunger | −100.0 |

In order to construct a labeled magnitude scale of satiety using these phrases, the magnitude estimates were first transformed by a multiplicative constant to range from +100 to −100. These values for the specific labels of the SLIM scale are shown in Table 3. A vertical, bi-directional visual analogue scale was then constructed with zero at the mid-point, +100 at the top and −100 at the bottom. Next to the zero point was placed the phrase 'neither hungry nor full'. The phrase 'greatest imaginable fullness' was placed at +100 and 'greatest imaginable hunger' was placed at −100. The other phrases were then placed at the appropriate distance along the top and bottom of the scale in accordance with their transformed magnitude estimates, producing the SLIM (satiety labeled intensity magnitude) scale shown in Fig. 1.

## Experiment 2

The purpose of the second experiment was to test the sensitivity and reliability of the newly developed SLIM scale and to compare it to visual analogue scales of hunger and/or fullness.

### *Method*

#### *Subjects*

Fifty subjects were randomly chosen from the same test P001 used in Experiment 1.

#### *Stimuli*

Image-based stimuli were used. Image based stimuli have been shown to produce similar data patterns to those of actual stimuli in a number of psychophysical applications (Baird & Harder, 2000) and have been shown to be a practical and efficient procedure for testing scale properties of validity, sensitivity and reliability (Cardello et al, 2003; Schutz & Cardello, 2001). The stimuli consisted of seven different written statements describing eating situations that

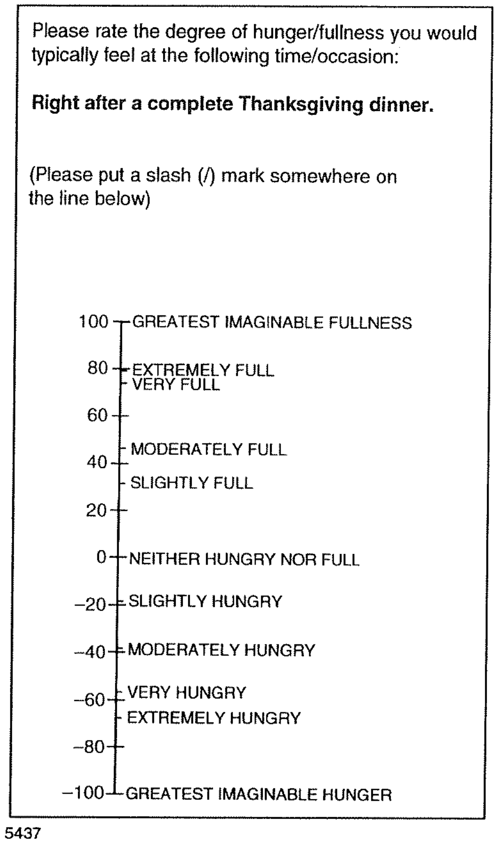


Fig. 1. The SLIM (Satiety Labeled Intensity Magnitude) scale.

logically evoke distinct cognitive levels of hunger/fullness. These written scenarios appear in the left hand column of Table 4. In addition, to evaluate the test–retest reliability of the scales, two of the scenarios were repeated, for a total of nine stimuli that were evaluated. Each written scenario appeared on a separate page of a questionnaire along with one of five scales used for assessing the satiety level evoked by the eating scenario.

*Scalar techniques*

Five different scales were assessed: (1) the 100 mm SLIM scale shown in Fig. 1, but without the scale numbers printed on the left side (previous research by Green et al. (1993) and Schutz & Cardello (2001) have confirmed that the presence of the numbers on labeled magnitude scales have little influence on subject responses), (2) a 100 mm, vertical, VAS hunger scale labeled 'not at all hungry' at the bottom and 'extremely hungry' at the top, (3) a 100 mm vertical, VAS fullness scale labeled 'not at all full' at the bottom and 'extremely full' at the top, (4) a 50 mm VAS hunger scale labeled the same at the top and bottom as the 100 mm hunger scale, and (5) a 50 mm VAS fullness scale labeled the same as the 100 mm fullness scales. The latter two scales were included to assess whether a shorter line length would produce better or worse sensitivity/reliability, since the SLIM scale is a 100 mm bi-directional scale with the hunger and fullness segments each equaling 50 mm in length.

*Procedure*

All testing was conducted in individual sensory testing booths. A repeated measures design was used, with the order of scales randomized across subjects. All nine written scenarios were evaluated, first using one scale, then the second scale, etc. The nine scenarios were placed together in a single questionnaire for each scale type. Written instructions were provided at the beginning of each questionnaire. Subjects made their ratings by placing a slash mark (/) somewhere through the line scale to indicate the degree of perceived satiety (hunger/fullness) evoked by

Table 4
Mean satiety ratings from Experiment 2

| | Mean response | | | | |
|---|---|---|---|---|---|
| | 100 mm Hunger scale | 100 mm Fullness scale | 100 mm Hunger scale | 100 mm Fullness scale | Slim scale |
| Right after a complete Thanksgiving dinner | 4.90 | 95.29 | 2.60 | 47.23 | 93.17 |
| Right after evening dinner with salad, entrée… | 6.52 | 82.17 | 5.63 | 41.15 | 84.40 |
| Right after evening dinner with salad, entrée… | 6.77 | 82.81 | 5.73 | 41.48 | 83.81 |
| Immediately after lunch | 12.10 | 66.90 | 10.48 | 34.85 | 72.15 |
| Two hours after lunch | 34.27 | 48.42 | 19.65 | 22.90 | 53.02 |
| Right before I normally eat lunch | 71.35 | 10.92 | 33.48 | 11.25 | 30.71 |
| After not having eaten anything for 24 h | 88.27 | 5.83 | 43.31 | 4.50 | 16.50 |
| After not having eaten anything for 24 h | 88.73 | 7.60 | 43.33 | 4.88 | 16.31 |
| After not having eaten anything for 3 days | 94.02 | 4.31 | 45.35 | 2.33 | 7.56 |

The values in parentheses next to the SLIM scale mean ratings are the means converted to a −100 to +100 scale.

the written scenario. After making their rating, subjects returned their questionnaire through a pass-through and awaited the next questionnaire.

*Results and discussion*

Data from all five scales were obtained by taking metric measurements along the line scales and arithmetic means of the ratings were calculated for all five scales.

Due to missing data, two subjects' data were dropped from consideration. The mean satiety ratings for each of the 9 scenarios (two replicated) for the remaining 48 subjects are shown in the right-hand columns of Table 4. As can be seen, the mean ratings for all five scales fall in the same as order the degrees of hunger/fullness logically expressed by the scenarios. In this regard, all five scales possess general construct validity. In addition, the mean ratings for the two replicated statements are similar on all five scales.

In order to assess the sensitivity of the different scales to the different levels of expressed satiety and to determine whether there was, in fact, no significant difference in mean ratings for the replicated statements, a one-way ANOVA and Tukey-HSD post-hoc tests were conducted for each scale type. The results of these analyses are shown in Table 5. The asterisks in Table 5 represent significant differences in means among the different pairs of statements. Optimal sensitivity is reflected by asterisks in all cells, with the exception of those two cells that represent the comparisons of the replicated statements (cell 2 v 8 and cell 6 vs 9). As can be seen in Table 5, only the SLIM scale showed perfect sensitivity. All other scales failed to show significant differences between one or more pairs of scenarios. In addition, an analysis of the pairs of scenarios for which these other scales failed to show discrimination indicates that 75% of them were between the two most extreme scenarios of hunger or fullness. These findings suggest that the advantage in sensitivity of the SLIM scale is greatest at the extremes of hunger and fullness, a fact that has been demonstrated in other studies of labeled magnitude scales (Schutz & Cardello, 2001; Cardello et al., 2003).

Pearson product–moment correlations were calculated between the ratings assigned by subjects to the two replicated statements for each scale type. Table 6 shows the results of these analyses. As can be seen, while there are differences in the correlation coefficients by scale and by replicated scenario, the average correlations are highest for the 100 mm visual analogue hunger scale (avg. $r=0.92$) and the SLIM scale (avg. $r=0.90$). It is not clear from the data whether a longer (100 mm) vs shorter (50 mm) visual analogue scale produces better reliability, since the data varied greatly between the two replicated statements. The same is true when one compares either hunger scale to either fullness scale.

Taken together, the results from this experiment indicate that the SLIM scale has somewhat better sensitivity to differences in expressed levels of hunger/fullness than any of the visual analogue scales tested and has test-retest reliability that is comparable or better than these scales.

## Experiment 3

In order to test the sensitivity and reliability of the SLIM scale using real food as stimuli, the following study was undertaken.

*Method*

*Subjects*

Study volunteers were 20 different employees of Natick Soldier Center drawn from the same pool of subjects as those used in Experiments 1 and 2. Subjects were asked to complete a background questionnaire to determine morning eating routines, so that test sessions could be scheduled to least disrupt subjects' normal eating habits.

*Stimuli*

Foods used in the study had been used in previous in-house studies of satiety (Merrill et al., 2002) and were found to be acceptable products. Based on this previous research and that of Holt et al. (1995), three foods that were found to vary greatly in perceived satiety were selected. The foods selected were strawberry yogurt (Axelrod), maple and brown sugar oatmeal (Quaker Instant), and croissants (Pillsbury Crescent). All foods were prepared according to manufacturer directions on the morning of the test and presented in 240 kcal portions in standard serving dishes.

*Scalar techniques*

In addition to the SLIM scale used in Experiment 2, two other satiety scales were used in order to assess the sensitivity and reliability of the SLIM scale relative to other VAS scales, but now, within the context of real foods. The first of these scales was the 100 mm, vertical unidirectional, VAS scale of hunger used in Experiment 2. This scale was included for comparison because it showed high reliability in Experiment 2. The second scale was a 100 mm, bi-directional hunger-fullness scale. This scale was chosen because it showed high reliability in previous research investigating the satiety value of foods (Merrill et al., 2002), but was not compared to the SLIM scale in Experiment 2.

*Procedure*

A within-subjects, repeated measures design was used. Subjects were instructed not to eat after 9:00 pm on the evening before each test session to establish an overnight fasting time of 10–12 h. On each of three mornings on two

Table 5
Results of Tukey-HSD post-hoc tests comparing the mean satiety values for all possible pairs of the nine eating scenarios (two replicates)

**100 MM HUNGER SCALE**

| Mean | Group | EAT_SIT7 | EAT_SIT9 | EAT_SIT6 | EAT_SIT5 | EAT_SIT4 | EAT_SIT3 | EAT_SIT2 | EAT_SIT8 | EAT_SIT1 |
|---|---|---|---|---|---|---|---|---|---|---|
| 4.90 | EAT_SIT7 | | | | | | | | | |
| 6.52 | EAT_SIT9 | | | | | | | | | |
| 6.77 | EAT_SIT6 | | | | | | | | | |
| 12.10 | EAT_SIT5 | | | | | | | | | |
| 34.27 | EAT_SIT4 | * | * | * | * | | | | | |
| 71.35 | EAT_SIT3 | * | * | * | * | * | | | | |
| 88.27 | EAT_SIT2 | * | * | * | * | * | * | | | |
| 88.73 | EAT_SIT8 | * | * | * | * | * | * | | | |
| 94.02 | EAT_SIT1 | * | * | * | * | * | * | | | |
| 27 SIGNIFICANT DIFFERENCES | | | | | | | | | | |

**50 MM HUNGER SCALE**

| Mean | Group | EAT_SIT7 | EAT_SIT9 | EAT_SIT6 | EAT_SIT5 | EAT_SIT4 | EAT_SIT3 | EAT_SIT2 | EAT_SIT8 | EAT_SIT1 |
|---|---|---|---|---|---|---|---|---|---|---|
| 2.60 | EAT_SIT7 | | | | | | | | | |
| 5.63 | EAT_SIT9 | | | | | | | | | |
| 5.73 | EAT_SIT6 | | | | | | | | | |
| 10.48 | EAT_SIT5 | * | | | | | | | | |
| 19.65 | EAT_SIT4 | * | * | * | * | | | | | |
| 33.48 | EAT_SIT3 | * | * | * | * | * | | | | |
| 43.31 | EAT_SIT2 | * | * | * | * | * | * | | | |
| 43.33 | EAT_SIT8 | * | * | * | * | * | * | | | |
| 45.35 | EAT_SIT1 | * | * | * | * | * | * | | | |
| 28 SIGNIFICANT DIFFERENCES | | | | | | | | | | |

**100 MM FULLNESS SCALE**

| Mean | Group | EAT_SIT1 | EAT_SIT2 | EAT_SIT8 | EAT_SIT3 | EAT_SIT4 | EAT_SIT5 | EAT_SIT9 | EAT_SIT6 | EAT_SIT7 |
|---|---|---|---|---|---|---|---|---|---|---|
| 4.31 | EAT_SIT1 | | | | | | | | | |
| 5.83 | EAT_SIT2 | | | | | | | | | |
| 7.60 | EAT_SIT8 | | | | | | | | | |
| 20.92 | EAT_SIT3 | * | * | * | | | | | | |
| 48.42 | EAT_SIT4 | * | * | * | * | | | | | |
| 66.90 | EAT_SIT5 | * | * | * | * | * | | | | |
| 82.17 | EAT_SIT9 | * | * | * | * | * | * | | | |
| 82.81 | EAT_SIT6 | * | * | * | * | * | * | | | |
| 95.29 | EAT_SIT7 | * | * | * | * | * | * | * | * | |
| 32 SIGNIFICANT DIFFERENCES | | | | | | | | | | |

**50 MM FULLNESS SCALE**

| Mean | Group | EAT_SIT1 | EAT_SIT2 | EAT_SIT8 | EAT_SIT3 | EAT_SIT4 | EAT_SIT5 | EAT_SIT9 | EAT_SIT6 | EAT_SIT7 |
|---|---|---|---|---|---|---|---|---|---|---|
| 2.33 | EAT_SIT1 | | | | | | | | | |
| 4.50 | EAT_SIT2 | | | | | | | | | |
| 4.88 | EAT_SIT8 | | | | | | | | | |
| 11.25 | EAT_SIT3 | * | * | * | | | | | | |
| 22.90 | EAT_SIT4 | * | * | * | * | | | | | |
| 34.85 | EAT_SIT5 | * | * | * | * | * | | | | |
| 41.15 | EAT_SIT9 | * | * | * | * | * | * | | | |
| 41.48 | EAT_SIT6 | * | * | * | * | * | * | | | |
| 47.23 | EAT_SIT7 | * | * | * | * | * | * | * | * | |
| 32 SIGNIFICANT DIFFERENCES | | | | | | | | | | |

**SLIM SCALE**

| Mean | Group | EAT_SIT1 | EAT_SIT8 | EAT_SIT2 | EAT_SIT3 | EAT_SIT4 | EAT_SIT5 | EAT_SIT6 | EAT_SIT9 | EAT_SIT7 |
|---|---|---|---|---|---|---|---|---|---|---|
| 7.56 | EAT_SIT1 | | | | | | | | | |
| 16.31 | EAT_SIT8 | * | | | | | | | | |
| 16.50 | EAT_SIT2 | * | | | | | | | | |
| 30.71 | EAT_SIT3 | * | * | * | | | | | | |
| 53.02 | EAT_SIT4 | * | * | * | * | | | | | |
| 72.15 | EAT_SIT5 | * | * | * | * | * | | | | |
| 83.81 | EAT_SIT6 | * | * | * | * | * | * | | | |
| 84.40 | EAT_SIT9 | * | * | * | * | * | * | | | |
| 93.17 | EAT_SIT7 | * | * | * | * | * | * | * | * | |
| 34 SIGNIFICANT DIFFERENCES | | | | | | | | | | |

KEY TO EATING SITUATIONS:
EAT_SIT1: AFTER NOT HAVING EATEN ANYTHING FOR 3 DAYS
EAT_SIT2: AFTER NOT HAVING EATEN ANYTHING FOR 24 HOURS
EAT_SIT3: RIGHT BEFORE I NORMALLY EAT LUNCH
EAT_SIT4: TWO HOURS AFTER LUNCH
EAT_SIT5: IMMEDIATELY AFTER LUNCH
EAT_SIT6: RIGHT AFTER EVENING DINNER WITH SALAD, ENTREE, ...
EAT_SIT7: RIGHT AFTER A COMPLETE THANKSGIVING DINNER
EAT_SIT8: AFTER NOT HAVING EATEN ANYTHING FOR 24 HOURS
EAT_SIT9: RIGHT AFTER EVENING DINNER WITH SALAD, ENTREE, ...

Asterisks in the cells indicate significant differences ($p < 0.5$) between the paired scenarios. (Legend on the lower right lists the scenarios corresponding to each 'Group number' in the table (e.g. EAT SIT8).

consecutive weeks, subjects reported to a research dining area at a scheduled time closest to their usual breakfast time (7:15–9:45 A.M.). Subjects were seated at a large dining table with dividers separating subjects so that they could not see or talk to one another. Upon arrival, subjects were asked to rate their baseline subjective hunger/fullness on each of the three test scales by placing a slash through the line to indicate their level of hunger/fullness. The order of scales was randomized among subjects. Participants were instructed to pay attention to the specific wording on each scale since they could be similar. Immediately after completing these baseline ratings, subjects were presented with one of the three test foods. No beverage was provided. Subjects were instructed to eat the entire sample of food at a comfortable rate but were asked to finish within 10 min. During any one test session, all subjects ate the same food, and each subject tested all three foods over the course of the three test sessions.

Table 6
Individual and average Pearson product-moment correlation coefficients calculated between ratings assigned to the two replicated scenarios for each scale type (Experiment 2)

| Response scale | Replicated scenarios | | |
|---|---|---|---|
| | Not eating for 24 h | Evening dinner | Average correlation |
| 100 mm Hunger | 0.95 | 0.89 | 0.92 |
| 100 mm Fullness | 0.83 | 0.32 | 0.58 |
| 50 mm Hunger | 0.98 | 0.55 | 0.77 |
| 50 mm Fullness | 0.54 | 0.93 | 0.74 |
| SLIM scale | 0.96 | 0.84 | 0.90 |

Immediately after consuming the food, subjects rated their liking/disliking of the food using the Labeled Affective Magnitude (LAM) scale (Schutz & Cardello, 2001). Following this and at every 10 min interval for the next 30 min, subjects rated their subjective hunger/fullness using the three test scales presented in random order. The thirty-minute test period was determined from the earlier work of Merrill et al. (2002), in which the correlation between the areas under the curve (AUC) for ratings obtained 30 and 60 min post-consumption were very high ($r=0.88$). These data suggested that a 30 min time interval is sufficient to ensure a representative level of post-consumption satiety, thereby enabling a meaningful analysis of the relative reliability and sensitivity of the different scales, while minimizing the total amount of time required to execute the study. Subjects were allowed to read, relax, or talk with one another between ratings, with the exception that they could not discuss the food or the study in any way. At the end of testing, a take-away, boxed continental breakfast was provided as a reward for participation and to provide immediate nutritional compensation for the fact that subjects could not eat anything from 9 PM the prior evening until the time of testing. The same subjects returned one week later, and the procedure was repeated with the same test foods in the same order to assess test-retest reliability of the scales. Subjects were not told the reason for returning, other than that their participation was required in another mealtime study.

*Results and discussion*

Fig. 2 shows the mean satiety responses obtained for the three foods and three scales at baseline and during the next 30 min. As can be seen, for all three foods, subjects had about the same degree of hunger/fullness at baseline. After consuming the food, subjects experienced an increase in fullness (reduction in hunger) for all three foods. In the case of yogurt there was a gradual reduction in fullness as time progressed, but for the other two foods, fullness remained relatively constant, with a small decrease seen only at 30 min. These results are to be expected, given the short duration of the testing. In general, the pattern of data for the three scales, is very similar over time. In fact, Pearson correlation coefficients calculated among the three scales over both time and food, showed the SLIM scale to be highly correlated with the bi-directional hunger-fullness scale (0.90), but less well correlated with the unidirectional hunger scale (−0.81). The correlation coefficient between the latter two scales was −0.80. These results are also to be expected, given that the SLIM scale is bi-directional, with 'hunger' and 'fullness' dimensions radiating in either direction from the zero-point. In this sense, the SLIM scale is more similar to the bi-directional hunger-fullness scale than is either to the unidirectional hunger scale.

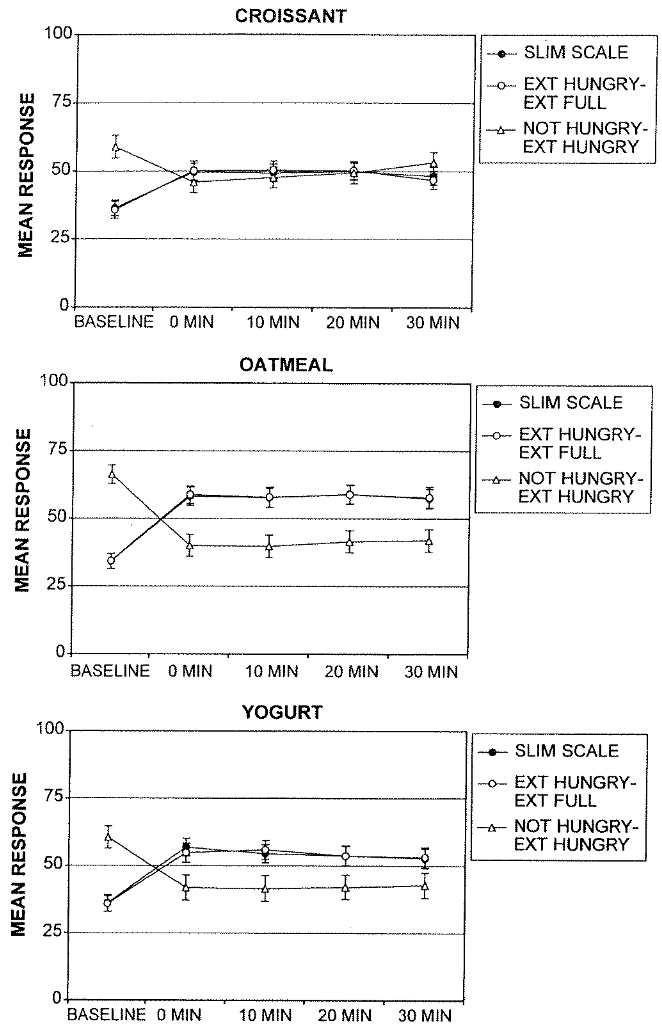


Fig. 2. Mean satiety responses at baseline and during first 30 min post-consumption for the foods and scales tested in Experiment 3.

Although all three scales performed somewhat similarly in terms of the mean ratings of satiety over time, the test-retest correlation coefficients for the three scales, calculated across time and foods, showed the SLIM scale to have the highest reliability ($r=0.54$) between ratings obtained on the two different weeks of testing. Its reliability was significantly higher ($p=0.05$) than that for the unidirectional hunger scale ($r=0.39$), but not significantly higher ($p=0.13$) than that for the bi-directional hunger-fullness scale ($r=0.43$).

## General discussion

The experiments reported here detail the development of an alternative scale for the assessment of satiety; one that is modeled after the labeled magnitude scales of Green et al. (1993) and Schutz and Cardello (2001). While the latter scales were developed for the purpose of scaling perceived oral sensation intensity and liking/disliking, the present Satiety Labeled Intensity Magnitude (SLIM) scale was developed specifically for the purpose of scaling hunger/fullness. A major advantage of this scale is that it produces data with ratio properties, enabling statements to be made about sensations of satiety being twice, one-half, etc. as intense as other sensations. This capability enables product and menu developers, for example, to be able to state that product A is twice (one-third, etc.) as filling as product B for the purpose of tailoring foods/rations to meet desired levels of perceived satiety. In subsequent research (Merrill et al., 2003) the SLIM scale has been used to index the perceived satiety of a range of ration items, so that ration developers can better assess the feasibility of combining items of low satiety value but high caloric density to create meals that increase total caloric intake. Since the SLIM scale is bi-directional, statements about the ratios among stimuli are restricted to sensations that all reflect hunger and/or to those that all reflect fullness (it is not meaningful to state that a sensation of fullness is 2× or 3× greater than a sensation of hunger). However, for those investigators interested in using the SLIM scale as two separate unidirectional scales (one for hunger, one for fullness), the two ends of the scale can be 'unfolded' and used separately. The two unfolded ends can each be easily re-constructed for use on paper or computerized ballots by utilizing the scale point locations for the SLIM labels shown in Table 3.

While the SLIM scale enables statements to be made about the ratios of perceived hunger (or fullness) associated with different foods/experiences, it avoids both the problems of unequal intervals and under-use of end-points that characterize category scales of satiety. The under-use of end-points necessarily limits sensitivity at the extremes of the scale. VAS scales can improve sensitivity at the extremes of the scale, but the results of Experiment 2 showed the SLIM scale to have even greater sensitivity than the four unidirectional VAS scales that were tested, especially at the highest levels of expressed hunger and fullness. This is a logical consequence of the SLIM scale end-points ('greatest imaginable hunger (fullness)') that enable more extreme ratings than merely 'extremely hungry (full)'. Thus, these end-point labels serve not only to help anchor different subject ratings to a common scale, but also to foster better discrimination of extreme levels of hunger/fullness. This can be an important advantage, because in many food intake studies, subjects are provided large meals to consume or are restricted in their intake of food for long periods of time.

It is worthwhile to point out that the SLIM scale was developed with the end-point labels of 'greatest imaginable hunger' and 'greatest imaginable fullness'. In her research on psychophysical scaling of taste intensity (Bartoshuk, 2000; Bartoshuk et al., 2001; Duffy et al., 1999), Bartoshuk uses the end-point label of 'greatest imaginable sensation *of any kind*'. Bartoshuk uses this end-point label in order to compare the responses of different individuals, some of whom have known differences in their experiences of taste intensity, e.g. supertasters. Of course, in the case of taste and many other sensory modalities, the adequate stimulus is well established, so it is relatively easy to determine that sub-groups in the population respond differentially to the adequate stimulus for that sensory quality. In the case of satiety, the adequate stimulus is much less well defined, and satiety appears to be multi-determined. While it is reasonable to speculate that the sensory *experience* associated with maximal fullness (or hunger) may be different for different sub-groups of the population, e.g. obese vs non-obese, there is no well documented psychophysical evidence showing this. More importantly, there is no evidence showing that their conceptions of the greatest *imaginable* fullness are any different. As Bartoshuk et al. (2004) has pointed out, the absolute intensity that one associates when using the term 'imaginable' (e.g. greatest imaginable sensation of any kind) may, in fact, be quite different from the absolute intensity associated with the greatest 'experienced' sensation: 'Imaginable may have a systematic association with the strongest experienced sensation, but it might also be open to a variety of interpretations' (Bartoshuk et al., 2004). Thus, while the greatest experienced sensation of satiety may differ among different sub-groups, it could be that the greatest *imaginable* satiety does not differ greatly. Or it may differ idiosyncratically, but not differentially, between obese and normal weight individuals.

The major reason for using the terms 'greatest imaginable fullness (hunger)' as end-points for the SLIM scale was to avoid a potential loss of sensitivity produced by using the much larger perceptual context of 'greatest imaginable sensation *of any kind*'. Contextual range effects on psychophysical judgments have been well demonstrated, and they occur across a wide range of stimulus-response modes (Helsen, 1964; Parducci, 1963, 1974; Poulton, 1989; Teghtsoonian and Teghtsoonian, 1978). One of the more robust findings from this research is that sensitivity to differences between sensations decreases when the rating context (range of endpoint anchors) is large. That is, when two stimuli are judged within a finite set of end-points, the stimuli are more difficult to discriminate when the end-points define a large perceptual range around the stimuli, as compared to when the end-points define a small range around them. The reason for this is that extreme anchors force the perceptions/ratings away from the anchors and toward the middle of the scale.

For example, take the situation in which two stimuli (300 and 400 kcal preload) produce two distinct satiety sensations that must be discriminated using a line of fixed length with extreme end-point labels, e.g. from 'no sensation' to 'greatest imaginable sensation of any kind'. If one were to identify a place on that line corresponding to some very intense, but non-satiety related sensation, e.g. the 'greatest imaginable loudness', it would likely be located well below the top of the line (assuming that pain sensations are, for most people, among the greatest of imaginable sensations and probably define 'greatest imaginable sensation of any kind'). Similarly, the 'greatest imaginable fullness' might be located, in all likelihood, still lower on the line (assuming that the greatest imaginable sensation resulting from an eating experience is less salient and intense than, let's say, 'being next to a ship's foghorn when it sounds'). The necessity of allowing, psychologically, for the placement on the scale of these other imaginably more intense sensations leaves a much restricted portion of the line in which to differentially rate the intensity judgments corresponding to the fullness produced by the 300 and 400 kcal stimuli. This 'scalar compression' would make discrimination of small differences more difficult. Although it is arguable whether these contextual effects occur in the perceptual domain or in the response domain, the impact on the data is the same.

Since the SLIM scale was developed for the primary purpose of measuring perceived satiety responses over time within the same individual and for differentiating the perceived satiety responses to different foods within a randomly selected heterogeneous group, as was the case in this study, the likely improvement in sensitivity produced by a considerably narrower contextual range was deemed more beneficial than using a much larger range, which would insure against only a future and *potential* demonstration that the term(s) 'greatest *imaginable* hunger (fullness)' differ systematically among different sub-groups in the population. Of course, if such differences are demonstrated, the scale construction task for the SLIM scale should be repeated to include 'greatest imaginable sensation of any kind'.

As to the reliability of the SLIM scale, the present studies show it to have equal or better reliability than most of the VAS scales studied here (Experiments 2 and 3). This better reliability may be attributable to the fact that the SLIM scale uses multiple verbal anchors, so that subjects can more precisely make their ratings along the horizontal line scale and can better reproduce those ratings when a similar sensation is experienced.

Other practical aspects of the SLIM scale that bear consideration include the fact that it uses simple verbal labels of hunger/fullness. Second is the fact that the specific numerical labels that appear on the scale are somewhat arbitrary. It appears from other studies conducted on labeled magnitude scales that subjects pay relatively little attention to the numbers on the scale (Green et al., 1993; Schutz & Cardello, 2001). It may well be the case that no numbers is a viable option in certain cases. This is particularly true if the data from the scale are to be compared among users who differ significantly in their knowledge or use of numbers (children vs adults) or where cultural or practical concerns may make the numbers more of a distraction. When no numbers are used, data from the scale can simply be transcribed from measurements made with a ruler on the 100 mm line and then transformed to a $-100$ to $+100$ scale. However, if it is desired to make ratio statements about hunger and or fullness, the scale must conform to the numerical values that were originally used to scale the semantic labels, i.e. $-100$ to $+100$, or to a multiplicative transformation of these values. A third consideration is that the graphic, continuous nature of the scale makes it amenable to computer-based, analogue to digital, and manual based encoding of data, and enables presentation in either a vertical or horizontal mode. The last practical aspect of the scale is that a simple arithmetic mean can be used as a measure of central tendency. This stands in contrast to magnitude estimation, where median, geometric means, or log transformations of the data must be calculated to arrive at a measure of central tendency.

The SLIM scale presented here has been used subsequently in studies in our laboratory to determine the perceived satiety value of rations (Merrill et al., 2003). These studies have shown the SLIM scale to be a simple and efficient technique for assessing the hunger/fullness associated with eating equi-caloric portions of different foods. These studies have also confirmed the greater sensitivity of the SLIM scale over a simple VAS scale of hunger-fullness for differentiating the satiety produced by different foods.

In summary, the SLIM scale developed here is a simple, easy to use instrument for assessing the satiety experienced by subjects in response to foods and/or other simulated eating situations (Experiment 2). The SLIM scale has good reliability and better discriminative sensitivity than other visual analogue scales of satiety, especially at the extremes. In addition it provides data similar to magnitude estimation, allowing ratio statements to be made about the relative hunger or fullness experienced in response to both real and imagined eating scenarios.

## Acknowledgements

This review contributes to a second Special Section on U.S. Army Research, originally conceived by H.L. Meiselman, F.M. Kramer, J. Dwyer, and D. Booth, and guest-edited by P. Pliner, H.L. Meiselman, and F.M. Kramer. The first Special Section on US Army Research appeared in 2003, Volume 40, Number 3.

## Appendix A. Instructions for modulus-free magnitude estimation (Experiment 1)

*Instructions*

In this test we would like to obtain your opinion about the meaning of different words and phrases that are commonly used to describe how HUNGRY or FULL one feels. In order to obtain your opinions about these words and phrases, we are going to use a special method that allows you to indicate the magnitude of hunger or fullness that is implied by each word or phrase by simply assigning numbers to represent your perception of their meaning.

On each of the pages that follow, you will find a word or phase that can be used to describe some degree of hunger or fullness. Next to it will appear two blank lines as in the following example:

| | +, -, 0 | How Much |
|---|---|---|
| Slightly Hungry | ________ | ________ |

After reading the word or phrase, the first thing that you must do is to tell us whether it reflects some degree of HUNGER or FULLNESS. If you feel it expresses some degree of HUNGER, you should place a negative sign (−) in the first column. If instead, you feel the word or phrase expresses some degree of FULLNESS, then you should place a positive sign (+) in the first column. If you feel that the word or phrase does not express EITHER hunger or fullness, then you should place a zero (0) in the first column.

After having determined whether the phrase expresses hunger, fullness, or neither and writing the appropriate symbol (−, +, 0) on the first line, you will then indicate the MAGNITUDE of the hunger *or* fullness reflected by the phrase. You will do this by placing a number on the second blank line (under 'How Much'). For the first phrase that you rate, you can write ANY number that you want on the line. We suggest you do not use a small number for this first word/phrase, because subsequent words/phrases may reflect much lower levels of hunger or fullness. Aside from this restriction, you can use any number you want.

For each SUBSEQUENT word/phrase, your numerical judgment should be made PROPORTIONALLY AND IN COMPARISON TO THE NUMBER YOU GAVE TO THE FIRST WORD/PHRASE. For example, if you assigned the number *100* to indicate the strength of the hunger/fullness reflected by the *first* word or phrase and the magnitude of the hunger or fullness of the second word/phrase is *twice* as great, then you would assign it a number twice as large, i.e. *200*. If it were *three* times as great, you would assign it the number *300*, etc. Similarly, if the second word or phrase denotes only *1/10* the magnitude of hunger or fullness as the first, you would assign it the number *10*, and so forth. If you judged any word or phrase to be 'neutral' (reflects neither hunger nor fullness, i.e. zero (0) in the first column), it should also be given a zero for its magnitude rating in the second column.

*Remember*. Proceed through each word or phrase by first judging whether it reflects HUNGER (−), FULLNESS (+), or neither (0). Then rate the strength of hunger or fullness reflected by the word/phrase by assigning a NUMBER to it that stands in the same *ratio* to the number assigned to the *first* word/phrase as does the *ratio* of the MAGNITUDE OF HUNGER/FULLNESS expressed by the second word phrase to that of the *first* word/phrase.

If you have any questions, please ask them before you begin. Thank you.

## References


Baird, J. C., & Harder, K. A. (2000). The psychophysics of imagery. *Perception and Psychophysics*, *62*, 113–126.

Barkeling, B., Rossner, S., & Sjoberg, A. (1995). Methodological studies on single meal food intake characteristics in normal weight and obese men and women. *International Journal of Obesity*, *19*, 284–290.

Bartoshuk, L. (2000). Comparing sensory experiences across individuals: recent psychophysical advances illuminate genetic variation in taste perception. *Chemical Senses*, *25*, 447–460.

Bartoshuk, L., Duffy, V. B., Chapo, A. K., Fast, K., Yiee, J. H., Hoffman, H. J., Ko, C., & Synder, D. J. (2004). From psychophysics to the clinic: missteps and advances. *Food Quality and Preference*. *15*, 617–632.

Bartoshuk, L., Duffy, V. B., Fast, K., Green, B., & Prutkin, J. (2001). Invalid sensory comparisons across groups: examples from PROP research. *Chemical Senses*, *26*, 761.

Blundell, J. E., & Burley, V. J. (1987). Satiation, satiety and the action of fibre on food intake. *International Journal of Obesity*, *11*(Suppl. 1), 9–25.

Booth, D. A. (1981). How should questions about satiation be asked? *Appetite: Journal of Intake Research*, *2*, 237–244.

Borg, G. (1982). A category scale with ratio properties for intermodal and interindividual comparisons. In H.-G. Geissler, & P. Petxoid (Eds.), *Psychophysical judgement and the process of perception* (pp. 25–34). Berlin: VEB Deutxcher Veriag der Wissenschaften, 25–34.

Burley, V. J., Leeds, A. R., & Blundell, J. E. (1987). The effect of high and low-fibre breakfasts on hunger, satiety, and food intake in a subsequent meal. *International Journal of Obesity*, *11*(Suppl. 1), 87–93.

Cardello, A. V., Winterhalter, C., & Schutz, H. G. (2003). Predicting the handle and comfort of military clothing fabrics from sensory and instrumental data: developmment and application of new psychophysical methods. *Textile Research Journal*, *73*, 221–237.

de Graaf, C. (1993). The validity of appetite ratings. *Appetite*, *21*, 156–160.

Duffy, V. B., Fast, K., Cohen, Z., Chodos, E., & Bartoshuk, L. M. (1999). Genetic taste status associates with fat food acceptance and body mass index in adults. *Chemical Senses*, *24*, 545–546.

Flint, A., Raben, A., Blundell, J. E., & Astrup, A. (2000). Reproducibility, power and validity of visual analogue scales in assessment of appetite sensations in single test meal studies. *International Journal of Obesity*, *24*, 38–48.

Green, S. M., & Blundell, J. E. (1996). Subjective and objective indices of the satiating effect of foods. Can people predict how filling a food will be? *European Journal of Clinical Nutrition*, *50*, 798–806.

Green, B. G., Shaffer, G. S., & Gilmore, M. M. (1993). Derivation and evaluation of a semantic scale of oral sensation magnitude with apparent ratio properties. *Chemical Senses*, *18*, 683–702.

Green, B. G., Dalton, P., Cowart, B., Shaffer, G., Rankin, K., & Higgins, J. (1996). Evaluating the 'labeled magnitude scale' for measuring sensations of taste and smell. *Chemical Senses*, *21*, 323–335.

Helsen, H. (1964). *Adaptation-level theory: An experimental and systematic approach to behavior*. New York: Academic Press.

Herman, C. P., Ostovich, J. M., & Polivy, J. (1999). Effects of attentional focus on subjective hunger ratings. *Appetite*, *33*, 181–191.

Hill, A. J., Magson, L. D., & Blundell, J. E. (1984). Hunger and palatability: tracking ratings of subjective experience before, during and after the consumption of preferred and less preferred food. *Appetite*, *5*, 361–371.

Hill, A. J., Rogers, P. J., & Blundell, J. E. (1995). Techniques for the experimental measurement of human eating behaviour and food intake: a practical guide. *International Journal of Obesity*, *19*, 361–375.

Hirsch, E. (1995). The effects of ration modifications on energy intake, body weight change, and food acceptance. In B. Marriot (Ed.), *Not eating enough: overcoming underconsumption of military operational rations* (pp. 151–173). Washington, DC: National Academy Press, 151–173.

Hirsch, E., & Kramer, F. M. (1993). Situational influences on food intake. In B. Marriott (Ed.), *Nutritional needs in hot environments* (pp. 215–243). Washington, DC: National Academy Press, 215–243.

Holt, S. A., Miller, J. C., Petoez, P., & Farmakalidis, E. (1995). A satiety index of common foods. *European Journal of Clinical Nutrition*, *49*, 675–690.

Holt, S. A., Miller, J. C., & Stitt, P. A. (2001). The effects of equal-energy portions of different breads on blood glucose levels, feelings of fullness and subsequent food intake. *Journal of the American Dietetic Association*, *101*, 767–773.

Hulshof, T., de Graff, C., & Westrate, J. A. (1993). The effects of preloads varying in physical state and fat content on satiety and energy intake. *Appetite*, *21*, 273–286.

Kissileff, H. R. (1981). Getting satiety taped: comments on Teghtsoonian, Becker and Edelman. *Appetite: Journal for Intake Reasearch*, *2*, 235–236.

Kissileff, H. R., Gruss, L. P., Thornton, J., & Jordan, H. A. (1984). The satiating efficiency of foods. *Physiology and Behavior*, *32*, 319–332.

Lane, H. L., Catania, A. C., & Stevens, S. S. (1961). Voice level: Autophonic scale, perceived loudness, and effect of side tone. *Journal of the Acoustical Society of America*, *33*, 160–167.

Lawless, H. T. (1977). The pleasantness of mixtures in taste and olfaction. *Sensory Processes*, *1*, 227–237.

Meiselman, H. L. (1995). Introduction to the concept and issues underlying underconsumption in military settings. In B. Marriot (Ed.), *Not eating enough: Overcoming underconsumption of military operational rations* (pp. 57–64). Washington, DC: National Academy Press, 57–64.

Meiselman, H. L., Hirsch, E. S., & Popper, R. D. (1988). Sensory, hedonic and situational factors in food acceptance and consumption. In D. M. H. Thomson (Ed.), *Food acceptability* (pp. 77–87). London: Elsevier, 77–87.

Merrill, E. P., Cardello, A. V., Kramer, F. M., Lesher, L. L., Schutz, H., & Kinchla, M., (2003). The development of a satiety index for rations. Paper presented at the Fifth Pangborn Sensory Science Symposium, Boston, MA.

Merrill, E. P., Kramer, F. M., Cardello, A. V., & Schutz, H. G. (2002). A comparison of satiety measures. *Appetite*, *39*, 181–183.

Moskowitz, H. R. (1977). Magnitude estimation: notes on what, how, when and why to use it. *Journal of Food Quality*, *1*, 195–227.

Parducci, A. (1963). Range-frequency compromise in judgment. *Psychological Monographs*, *2* (Whole 565).

Parducci, A. (1974). Contextual effects: A range-frequency analysis. In E. C. Carterette, & M. P. Friedman, *Handbook of perception. Psychophysical judgment and measurement* (Vol. II) (pp. 127–141). New York: Academic Press, 127–141.

Porrini, M., Croveti, R., Tesolin, G., & Silva, S. (1995). Evaluation of satiety sensations and food intake after preloads. *Appetite*, *25*, 17–30.

Poulton, E. C. (1989). *Bias in quantifying judgments*. Hillsdale, NJ: Erlbaum.

Prutkin, J., Duffy, V. B., Etter, L., Fast, K., Gardner, E., Lucchina, L., Snyder, D., Tie, K., Weiffenbach, J., & Bartoshuk, L. M. (2000). Genetic variation and inferences about perceived taste intensity in mice and men. *Physiology and Behavior*, *69*, 161–173.

Robinson, R. G., Folstein, M. F., & McHugh, P. R. (1978). Changes in appetite and food intake following small bowel bypass surgery for obesity. *International Journal of Obesity*, *2*, 489–490.

Robinson, R. G., McHugh, P. R., & Folstein, M. F. (1975). Measurement of appetite disturbances in psychiatric disorders. *Journal of Psychiatric Research*, *12*, 59–68.

Rolls, B. J., Bell, E. A., Castellanos, V. H., Chow, M., Pelkman, C. L., & Thorwart, M. L. (1999). Energy density but not fat content of foods affected energy intake in lean and obese women. *American Journal of Clinical Nutrition*, *69*, 863–871.

Rolls, B. J., Kim-Harris, S., Fischman, M. W., Foltin, R. W., Moran, T. H., & Stoner, S. A. (1994). Satiety after preloads with different amounts of fat and carbohydrate: implications for obesity. *American Journal of Clinical Nutrition*, *60*, 476–487.

Schutz, H., & Cardello, A. V. (2001). A labeled affective magnitude (LAM) scale for assessing food liking/disliking. *Journal of Sensory Studies*, *16*, 117–159.

Silverstone, J. T., (1966). The measurement of hunger in relation to food intake. *Proceedings of the VII International Congress on Nutrition, Hamburg* (pp. 51–53).

Silverstone, J. T., & Stunkard, A. J. (1968). The anorectic effect of dexamphetamine sulphate. *British Journal of Pharmacology and Chemotherapy*, *33*, 513–522.

Stevens, J. C., (1957). *A comparison of ratio scales for the loudness of white noise and the brightness of white light*. Doctoral Dissertation. Harvard University, Cambridge, MA.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, *64*, 153–181.

Stevens, S. S. (1962). The surprising simplicity of sensory metrics. *American Psychologist*, *17*, 29–39.

Stevens, S. S. (1971). Issues in psychophysical measurement. *Psychology Review*, *78*, 426–450.

Stevens, S. S., & Galanter, E. H. (1957). Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology*, *54*, 377–411.

Stubbs, R. J., Hughes, D. A., Johnstone, A. M., Rowley, E., Reid, C., Elia, M., Stratton, R., Delargy, H., King, N., & Blundell, J. E. (2000). The use of visual analogue scales to assess motivation to eat in human subjects: a review of their reliability and validity with an evaluation of new hand-held computerized systems for temporal tracking of appetite ratings. *British Journal of Nutrition*, *84*, 405–415.

Stubbs, R. J., Johnstone, A. M., O'Reilly, L. M., & Poppitt, S. D. (1998). Methodological issues relating to the measurement of food, energy and nutrient intake in human laboratory-based studies. *Proceedings of the Nutrition Society*, *57*, 357–372.

Teghtsoonian, M., Becker, E., & Edelman, B. (1981). A psychophysical analysis of perceived satiety: its relationship to consumatory behavior and degree of overweight. *Appetite*, *2*, 217–229.

Teghtsoonian, R., & Teghtsoonian, M. (1978). Range and regression effects in magnitude scaling. *Perception and Psychophysics*, *24*, 305–314.

Wooley, S. C., & Wooley, O. W. (1973). Salivation to the sight and thought of food: a new measure of appetite. *Psychosomatic Medicine*, *35*, 136–142.

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.