ELSEVIER

Food Quality and Preference 14 (2002) 165–176

Food Quality and Preference

www.elsevier.com/locate/foodqual

Repeatability in hedonic sensory measurement: a conceptual exploration

E.P. Köster*, T. Couronne, F. Léon, C. Lévy, A.S. Marcelino

Centre Européen des Sciences du Goût, Dijon, France

Accepted 23 June 2002

Abstract

The traditional view of measurement repeatability is discussed in the light of psychological theories about stability and change in preference and choice behavior. The argumentation is illustrated by data obtained in groups of children and adults who are exposed to the same hedonic sensory measurements a number of times. It is demonstrated that first hedonic impressions are poor predictors of final liking and choice. The repeatability of hedonic methods should be judged on the basis of the stability of the change in preference of different, but comparable populations, rather than on the reliability of repeated measurement in the same population.

Keywords: Repeatability; Food choice; Preference

1. Introduction

In judging the effectiveness of sensory methods, three criteria are usually considered: the sensitivity or discrimination power of the method, the reliability or repeatability of the method and its external validity. The last of these three criteria is often neglected or, even worse, taken for granted. The first one is often the subject of studies in which two or more methods are compared on the basis of the number of significant differences between the same set of stimuli found with each of them. Most emphasis is usually laid on the measurement of the reliability of the method. When repeated with the same group and the same stimuli, the method should give the same results. This criterion which is essential when dealing with the measurements of events in physics or chemistry, is rather questionable when dealing with human decisions in general and especially when dealing with hedonic responses. Even in pure psychophysics, where the perceived intensity of stimuli is measured, the intrinsic variation of human sensory sensitivity poses many problems. Individual psychophysical functions may vary quite considerably from day to day even if the averaged group results stay the same (Cain cited in Engen, 1971; Punter, 1976; Punter & Köster, 1977). Adaptation and habituation play a very important role as safeguards against monotonous stimulation in human perception and especially in olfaction. Also, in hedonic decisions about likes and dislikes for a product many other factors, related to the changing internal state of the perceiving subject, add to the doubtfulness of repeatability as a good criterion to measure the effectiveness of methods. The effects of loss of initial curiosity, of arising product boredom and of slowly rising product irritation are at the base of many failures in the prediction of market success by methods that rely on first judgment (Köster, 1990, 1991).

As authors from different disciplines (Cabanac, 1971; Rolls, Rowe, Rolls, Kingston, Megson, & Gunary, 1981; Van Trijp, 1994) have indicated, there is a true search for variation in our appreciation of foods. Usually we do not eat the same food every day, when the possibility of variation exists. And even within a meal we like variation. Certainly, rice, potatoes, bread or pasta are eaten every day in different parts of the world, but they are consumed with different side dishes on different days. It seems that repeatability in our food choices is an exception rather than a rule.

The explanations for this phenomenon vary from alliesthesia, the idea that, due to metabolic mechanisms satiety for a given nutrient diminishes the appetite for it and may raise the appetite for another nutrient (Cabanac) and specific sensory satiety, the idea that these changes in appetite are due to sensory rather than to metabolic mechanisms (Rolls) to variety seeking, the idea that people differ in their psychological need for variation (Van Trijp). Although there is physiological

* Corresponding author. Present address: Jan van Scorelstraat 55, 3583 CK Utrecht, The Netherlands. Tel.: +31-30-251-0387; fax: +31-30-254-6071.

E-mail address: ep.koster@wxs.nl (E.P. Köster).

0950-3293/02/$ - see front matter
PII: S0950-3293(02)00075-7

evidence for the first two theories, the third one remains very near to the purely descriptive level. Furthermore, it is often hard to account for differences in the extent to which different foods or meal components are affected by these factors. Taking variety seeking as a personality trait (Van Trijp, 1994) seems too unspecific and does not explain much of actual choice behavior (Låhtenmäkki & van Trijp, 1995). On the other hand, taking variety seeking as the expression of specific satiety for certain nutrients (Cabanac, 1971) seems too specific and does certainly not explain all changes in preference.

In other sense modalities, psychologists have been preoccupied with variety seeking and changes of preference for many years. Since Tolman (1925) discovered that rats who entered the same T-maze for a second time would always take the arm they had not visited before, Spontaneous Alternation Behavior (SAB) has been the subject of many theoretical considerations. Glanzer (1953), using a T-maze with arms that differed in visual stimulation, showed that SAB was the result of a search for variety of stimulation and not of a tendency to vary locomotion patterns, as had been assumed till then. He ascribed this to a phenomenon he called stimulus satiation, a term borrowed from Gestalt psychology. Exposure to a given stimulus reduces the chance that it will be chosen on the next trial. Along the same lines, Dember (chap. 10, 1970) and Richman, Dember, and Kim (1987) propose that SAB may reflect the effort of the animal to increase environmental novelty for itself by exploring the arm of the T-maze least recently visited. Thus, SAB would also be a manifestation of exploratory behavior. This view reflects the theories of Berlyne (1970), who made a distinction between specific exploratory behavior and diversified exploratory behavior. In specific exploratory behavior, where the subject is in a conflictual state, due to a lack of information about the nature of a stimulus, curiosity (individual characteristic) and novelty (stimulus property) prevail, whereas in the diversified exploratory behavior, where the individual is no longer in this conflictual state, other stimulus properties like arousal potential and (perceived) stimulus complexity become predominant. The complexity is either defined a priori on the basis of information theoretical considerations (number of bits in a pattern) or is measured with the help of rating scales or paired comparison. Along the same lines as Berlyne, Young (1968) proposes the idea that "brief exposure tests imply a dynamic conflict between responses for which an organism is primed". This idea joins the response competition hypothesis forwarded by Harrison, Tutone, and McFadgen (1971): "when the perceiver is first confronted with novel stimulus, there are a number of antagonist response tendencies, producing a tension state. Subsequent exposure provides the opportunity for some of the response tendencies to be strengthened, while others are weakened or crowded out".

From still another angle, it should be mentioned that according to Dember and Earl (1957), Dember, Earl, and Paradise (1957) and Dember (chap. 10, 1970), preferences depend also on the individual ability to appreciate complexity (or novelty), which in turn is related to personal experience. When exposed to stimuli of slightly higher complexity than those that they appreciate most, people would have a tendency to raise their level of optimal complexity. This would result in a growing appreciation for these stimuli, which they called 'pacers', and a diminution of appreciation for the stimulus that hitherto the subjects liked most.

Finally, the effects of "mere exposure" Zajonc (1968) and neophobia have been described. Pliner (1982) using novel stimuli, showed that over repeated exposition the original fear for these stimuli was overcome, a phenomenon that she ascribed to the dissipation of neophobia.

Change of preference over repeated exposure in a free choice situation has also been observed in some human experiments unrelated to food. Brickman and d'Amato (1975) used a jukebox with eight unknown japanese songs in a free choice paradigm. Their subjects chose one of the songs on each of 40 trials. At first, the subjects satisfied their curiosity by a systematic exploration of the stimuli. Only in the later trials did repetitions of a given stimulus appear more frequently. Nevertheless, immediate repetitions remained seldom and in most of the repetitions other stimuli were allowed to intervene. Subjects showed both a preference for repetition (of the preferred stimulus), and a preference for variety. Table 1 gives an overview of the theories discussed earlier.

2. Experimental illustration

In fact, on the basis of this literature one must expect that first impressions are bad predictors of final liking. When novelty indeed dominates in the specific exploratory phase and this influence makes place for other characteristics later on, changes in preference will be largest for new and unfamiliar products. That this is the case is illustrated by the results of an experiment by F. Léon (1998; Léon, Couronne, Marcuz, & Köster, 1999) who presented five different jam-topped biscuits to 157 children between 5 and 10 years of age using two different methods (hedonic categorization and paired comparison) in two sessions each. With the paired comparison all children judged all 10 possible pairs once in each session. In order to equalize the stimulus exposure for both methods, in the hedonic judgement sessions each of the five stimuli were monadically presented four times in a random order.

A four point Smiley scale (dislike very much—dislike—like—like very much) was used. Three of the biscuits

Table 1
Overview of theories on change of preference and choice behavior

Theories	Phenomenon	Explaining principle	Verification
Food related theories			
Alliesthesia Cabanac (1971)	Change in liking for a food during eating	Satiety for specific nutrients transmitted by guttural receptors in stomach and intestine	Psychophysical and physiological measurements
Sensory specific satiety Rolls et al. (1981), Rolls (1999)	Change in liking for a food during eating	Satiety for specific sensory qualities due to stimulation of specific prefrontal cells	Hedonic scales and electro- physiology
Variety seeking Van Trijp (1994)	Change in liking for a food during eating and over meals	Psychological typology based on tendency to look for variety	Variety seeking scales
General psychological theories			
Spontaneous Alternating Behavior (SAB) Tolman (1925)	Rats choose different arms of a T-maze on successive trials	Need for change in locomotion patterns	Behavioral animal research
Stimulus satiation Glanzer (1953)	Rats choose differently colored arms of a T-maze on successive trials	Exposition to sensory stimuli satiates and causes need for perceptual change	Behavioral animal research
Novelty exploration Dember (1970) Richman et al. (1987)	Rats choose the arms of a T-maze least recently visited	Self stimulation by increasing environmental novelty	Behavioral animal research
Specific vs. diversified exploratory behavior Berlyne (1963) Young (1968), Harrison et al. (1971)	Preference for stimuli is different during the first (or first few) exposure(s) to them and during later exposures	Specific: Subject is in conflict state through lack of information: novelty and curiosity prevail Diversified: Subject no longer in conflict state: arousal potential and complexity (stimulus properties) prevail.	Human research on preference and aesthetic judgments
Optimal Arousal Level theory Berlyne (1970)	Different people prefer different stimuli, depending on the experience they have had and on the arousal they seek.	Each person has his/her own optimally preferred level of arousal. The same stimulus may have a higher arousal potential for the one person than for another	Human research on preference and aesthetic judgments
Developmental theories			
Mere Exposure theory Zajonc (1968), Pliner (1982)	Experience with stimuli leads to increased liking for them	Dissipation of Neophobia (fear for novel stimuli) leads to increasing liking	Repeated exposition to novel stimuli
Optimal Arousal theory Berlyne (1970)	Experience with stimuli leads to increased or to decreased liking (boredom)	Experience reduces the arousal potential of the stimuli and increases the subjects'optimal arousal level	Human research on aesthetics
"Pacer" theory Dember (1970), Dember et al. (1957), Dember and Earl (1957)	People learn to appreciate more complex stimuli in life (music, paintings, foods, wines, etc.)	Exposure to stimuli (called "pacers") that are slightly more complex than optimal leads to heightening of the subjects' optimal complexity (=arousal) level	Animal studies with visual patterns

(apricot, raspberry and strawberry) were popular in the french market, the other two (banana and lemon) were made especially for the experiment and therefore unknown to the children. On two of the biscuits (raspberry and strawberry) the color of the jam was red, on one of them (apricot) it was orange and on the other two (banana and lemon) it was yellow.

The results of this experiment have been described and analyzed in the traditional way elsewhere (Léon et al., 1999). Here, the data will be analyzed in a different way, paying attention only to the changes in preference. During this exercise, the reader is asked not to look immediately for statistical significance, but to see the text rather as a form of hypothesis formation and just to inspect the data and to follow the reasoning and the conclusions critically. For both methods the individually most preferred stimulus in each pair of stimuli has been determined for each subject in each session. In the case of hedonic judgments these pair preferences were constructed by averaging the four responses obtained for each stimulus and comparing the means of the two members of each of the possible pairs. In the case of ties, which occurred in less than 20% of the cases, the preference of the subject was divided equally over the two pair members. In paired comparison a forced choice procedure was used and as a consequence there were no ties. Subsequently, the percentage of the subjects who, over sessions, changed their preference from one pair member to the other was calculated for each possible pair of stimuli.

Table 2 shows the percentages of the children that changed their original preference for a given biscuit to another biscuit between the two sessions. The percentages change in each pair are given and the average percentages of these changes over all pairs containing a given originally preferred stimulus are given in the last column. It should be noted, that the numbers on which the percentages in the cells are based may differ since they depend on the number of children who showed an original preference for the stimulus mentioned on the left. To give an impression of the overall differences in liking for the different stimuli, the average ranks of the preference of the whole group over the two combined sessions are given with an indication of the statistical significance of the differences between them (P at least <0.05).

As can be seen from this table, with both methods the largest average changes (51.4 and 52.7% in Table 2a and 40.4 and 37.4% in Table 2b) are indeed found for the two novel products, that are also the least liked products (ranks between 3.80 and 3.95). At the same time it is clear that the changes in the different pairs are made predominantly towards the three better known and better liked (ranks between 2.09 and 2.61) products. However, with hedonic judgment (Table 2a) almost as high a percentage of the children, who originally chose lemon or banana, change to the other yellow and unfamiliar biscuit [lemon to banana (48.2%) and banana to lemon (48.7%)] as to strawberry (49.2 and 50.9%, respectively) and 43.4% of the children who originally preferred the most liked strawberry change to apricot, which is liked significantly less. Furthermore, it is remarkable that with hedonic categorization (Table 2a) all percentages change are higher than with paired comparison (Table 2b). Thus, it would seem that paired comparison is the more repeatable method. However when looking for an explanation of this difference in reliability, one may argue that in paired comparison visual information can play a much larger role than with the monadic presentation used in hedonic categorization.

Table 2
Biscuits with jams in natural colors

		Percentage change to:					
	Rank	Apricot	Banana	Lemon	Raspberry	Strawberry	Average
2a. Hedonic judgments							
Originally liked:							
Apricot	2.51b	–	28.6	28.5	53.8	53.7	37.7
Banana	3.92c	52.4	–	48.7	56.3	50.9	51.4
Lemon	3.95c	59.4	48.2	–	60.3	49.2	52.7
Raspberry	2.53b	42.4	29.3	27.0	–	48.8	34.6
Strawberry	2.09a	43.4	29.5	23.5	49.2	–	35.1
2b. Paired comparison							
Originally liked:							
Apricot	2.68b	–	15.5	16.4	35.8	33.0	27.4
Banana	3.80c	49.1	–	28.4	43.2	50.0	40.4
Lemon	3.90c	47.5	25.9	–	60.3	42.1	37.4
Raspberry	2.38ab	20.2	15.3	27.0	–	55.0	28.4
Strawberry	2.22a	43.4	23.1	13.7	34.2	–	23.7

$N=157$. Hedonic judgments (2a) and Paired comparison (2b) Percentages of children that exchange their originally most liked biscuit (session 1) for another most liked biscuit (session 2). Also given are the average percentage change away from the originally most liked biscuit (average) and the average rank of preference over both sessions for the group (rank 1 = most liked; rank 5 = least liked). Ranks with different letters are statistically different ($P<0.05$).

In paired comparison the differently colored stimuli are both directly in front of the child, whereas in the hedonic categorization the stimuli have to be compared with an internal reference. This was tested in an experiment (Léon, 1998) with 129 new children and with biscuits of the same shape and taste as in the previous experiment, but now having all the same red color. The results are given in Table 3.

As can be seen from these tables, the differences between the methods in percentage average change have almost disappeared and the data for the paired comparison fall also into the same range as those for the hedonic judgments with the differently colored biscuits (Table 2a). With hedonic judgment there are no major differences between the results of the experiments (Tables 2a and 3a). The only clear exception is the very high percentage of change (61.1%) found for red lemon in Table 3a. Obviously, red lemon does not fit the internal reference very well. Red banana does much better in this respect.

Another question that arises when one looks at the rather sizeable changes in the overall comparisons (37.3–61.1% in Table 3a and 41.0–53.2% in Table 3b), is whether these changes follow a random pattern or whether they show a direction of change. In the tables shown so far the latter is usually the case. Thus, in Tables 2a and 2b the changes of preference for those who preferred apricot in the first session went to a much larger degree to strawberry and raspberry than to banana and lemon. As indicated above, this might be due to the differences in familiarity with the stimuli or to differences in color. In the same way some of the unevenness of change distribution in Tables 3a and 3b might be due to the artificial coloring of otherwise well known tastes. How then would this be when only red fruits were used and some of them were familiar and others unfamiliar? In order to answer this question and to see the effects of some other variables like perceived complexity, an experiment with two familiar (strawberry and raspberry) and two unfamiliar (cherry and wild strawberry) red fruits was carried out. This time 180 new children took part. The wild strawberry taste was somewhat more complex and more pronounced than the ordinary strawberry taste and could be seen as a ''pacer'' in the sense of Earl and Dember, although independent proof for such a statement is lacking. The results of the experiment are given in Table 4.

As can be seen from this table, the values for change to wild strawberry are indeed highest in all rows with the exception of the first row of Table 4b where the change from cherry to raspberry is slightly more important. Furthermore, it is evident that, with the exception of raspberry, the percentage of children who change from a given original preference towards wild strawberry is larger than the percentage children who change from wild strawberry to that given biscuit. All of this is in accordance with the hypothesis that wild strawberry acts as a pacer. However, an alternative hypothesis might be that these changes were merely the result of the dissipation of neophobia. Wild strawberry was unknown till the first session. The children might therefore show a neophobic reaction to it, but after having been exposed to it three times during that session, this neophobia would be resolved and they liked the wild strawberry much better in the second session.

There are two reasons to reject this hypothesis. First of all, wild strawberry was already among the two most liked biscuits in the first session and secondly such improvement was not found for the other novel stimulus (cherry). Here, on the contrary, in all cases changes away from cherry were larger than changes towards it.

Table 3
Biscuits with all red jams in different flavors—hedonic judgments (3a) and paired comparison (3b)

		Percentage change to:					
	Rank	Apricot	Banana	Lemon	Raspberry	Strawberry	Average
3a. Hedonic judgments							
Originally liked:							
Apricot	2.59a	–	41.9	33.1	51.3	61.0	45.9
Banana	3.33b	53.2	–	30.9	57.9	54.1	47.4
Lemon	3.87c	57.1	63.3	–	55.4	67.7	61.1
Raspberry	2.62a	48.5	50.0	30.8	–	54.4	44.9
Strawberry	2.59a	51.1	35.8	27.4	40.4	–	37.3
3b. Paired comparison							
Originally liked:							
Apricot	2.77a	–	32.8	39.4	51.9	40.0	40.4
Banana	3.26b	51.2	–	40.3	53.3	55.1	49.2
Lemon	3.79c	40.5	54.2	–	65.8	51.6	53.2
Raspberry	2.55a	56.2	21.9	30.6	–	51.0	38.3
Strawberry	2.62a	53.3	41.0	30.4	42.6	–	41.0

Legend see Table 2. $N = 129$.

Table 4
Biscuits with all red jams—Hedonic judgments (4a) and paired comparison (4b)

		Percentage change to:				
	Rank	Cherry	Raspberry	Strawberry	Wild strawberry	Average
4a. Hedonic judgments						
Originally liked:						
Cherry	3.49a	–	40.0	58.0	68.2	56.2
Raspberry	2.83b	34.9	–	46.8	51.7	43.9
Strawberry	2.95b	40.0	55.9	–	63.8	52.3
Wild strawberry	2.91b	45.1	53.4	51.6	–	49.6
4b. Paired comparison						
Originally liked:						
Cherry	3.23a	–	52.7	43.3	49.2	48.5
Raspberry	3.10a	42.4	–	46.4	57.6	48.2
Strawberry	2.94ab	37.5	45.1	–	68.2	48.7
Wild strawberry	2.64a	44.3	42.0	38.6	–	41.7

$N = 180$. Legend see Table 2.

As its novelty disappears, it is liked less well instead of better as should be the case when the dissipation of neophobia was involved.

From all of this, it becomes clear that many changes in preference occur between the first and the second session, that these changes are not just due to random behavior and that they may contain important information.

Two further questions arise from these findings. Are two sessions enough to gain insight in the changes that occur? Do such changes also occur in adults or are they typical for children? In order to answer the first question two experiments with respectively, 170 and 199 new children were executed over three sessions. In the first experiment salted crackers and in the second one chocolate creams were used. Table 5 shows the changes in the hedonic judgments between session 1 and sessions 2 (Table 5a and 5d), between sessions 1 and 3 (Table 5b and 5e) and between sessions 2 and 3 (Table 5c and 5f), respectively.

Table 5
Crackers and chocolate creams

Hedonic judgment	Crackers ($N = 170$)						Chocolate creams ($N = 199$)					
5a.		Percentage change in session 2 to:					5d	Percentage change in session 2 to:				
Liked most in session 1	Rank	A	B	C	D	Average	Rank	A	B	C	D	Average
A	2.49a	–	24.0	33.3	30.4	28.9	2.21a	–	22.3	21.5	45.5	28.3
B	4.04d	67.4	–	47.2	53.9	55.5	2.43a	52.2	–	47.8	53.6	50.8
C	2.98b	47.1	33.3	–	35.6	37.7	3.86b	58.2	41.0	–	56.4	48.9
D	3.25c	48.8	33.2	46.4	–	41.5	4.24c	51.7	23.3	21.4	–	29.5
5b.		Percentage change in session 3 to:					5e	Percentage change in session 3 to:				
Liked most in session 1	Rank	A	B	C	D	Average	Rank	A	B	C	D	Average
A	2.49a	–	29.7	40.2	37.2	35.3	2.21a	–	22.1	25.5	46.5	29.8
B	4.04d	61.0	–	48.1	52.6	53.3	2.43a	57.9	–	67.2	59.2	62.4
C	2.98b	56.7	34.7	–	39.6	41.8	3.86b	58.8	33.8	–	48.7	42.9
D	3.25c	61.2	35.4	47.6	–	46.1	4.24c	49.3	22.7	23.3	–	29.2
5c.		Percentage change in session 3 to:					5f	Percentage change in session 3 to:				
Liked most in session 2	Rank	A	B	C	D	Average	Rank	A	B	C	D	Average
A	2.49a	–	20.6	36.0	30.9	28.5	2.21a	–	12.7	19.0	36.5	21.3
B	4.04d	42.4	–	45.0	48.8	45.3	2.43a	42.5	–	53.7	48.3	49.4
C	2.98b	49.6	25.5	–	33.7	34.5	3.86b	40.5	21.3	–	46.3	32.0
D	3.25c	47.9	31.8	35.7	–	37.1	4.24c	38.7	12.2	20.5	–	21.9

Changes from session 1 to session 2 (5a and 5d), from session 1 to 3 (5b and 5e) and from session 2 to 3 (5c and 5f). Legend see Table 2.

As can be seen from these tables the changes from session 1 to session 3 (Table 5b and 5e) are rather similar to these from session 1 to session 2 (Table 5a and 5d), but they are somewhat more pronounced, especially in the case of the changes of crackers C and D towards cracker A and in the case of the change of chocolate cream B to C. Only in a few cases (the change of B to A in the crackers and the changes of C to D and of D to both A and B in the chocolate creams) are the changes from session 1 to 3 smaller than those from session 1 to session 2. A comparison of the average changes in the top two tables (5a and 5d) on the one hand and those in the tables below them (5b and 5e) on the other shows a varied picture. In the transition from session 1 to 3, five of the eight averages are higher and 3 are lower than the corresponding averages in the transition from session 1 to session 2. Does this mean that the children who already changed their preference in session 2 stay with their second choice and are they just joined by some others or is everybody who moved before moving again and is there just some net gain for certain crackers or chocolate creams?

Inspection of the changes between session 2 and session 3 may already bring a partial answer. The results are given in Table 5c and 5f.

A comparison of the changes in Table 5c with those in Table 5a shows that for the crackers in all cases except the changes from product A to C and D and from product C to A, the changes between the sessions 2 and 3 are smaller than the ones between sessions 1 and 2. For the chocolate creams there is only one exception to this rule (the change from product B to C). This suggests a certain stabilization, although most amounts of change found between sessions 2 and 3 remain in the range from 30 to 50% for crackers and in a range of 10–50% for chocolate creams. Nevertheless, it is clear that the changes from session 1 to session 3 are not a simple summation of those between session 1 and 2 and those between session 2 and 3.

A more direct way of measuring the changes in the preference behavior of the children over the three sessions might be to look at the individual behavior of each child and to see how many children show the same behavior. In fact, for each pair of stimuli, there are only four types of change behavior possible. The child may stick to the same most preferred stimulus over all sessions (S–S), it may stick to the same preference between sessions 1 and 2, but change between sessions 2 and 3 (S–C), it may change between sessions 1 and 2, but remain stable between sessions 2 and 3 (C–S) or it may change between all successive sessions (C–C). In the latter case, the child returns to its original preference after having chosen the other alternative during the second session. For each of the two methods used in the experiments with the salted crackers and the chocolate creams, Table 6 shows the frequencies with which these four alternative behaviors occur for each pair of products, but irrespective of the original preference.

Table 6
Types of change behavior, irrespective of original preference for each of the stimulus pairs

6a. Crackers

Hedonic judgments	Δ Rank	S–S	S–C	C–S	C–C
Pair:					
A–B	1.55	42.8	15.0	24.3	17.9
A–C	0.49	35.3	19.6	17.9	27.2
A–D	0.76	37.0	19.1	20.2	23.7
B–C	1.06	39.3	16.8	22.0	22.0
B–D	0.79	37.0	16.8	19.1	27.2
C–D	0.27	37.0	16.8	22.0	24.3
Hedonic judgments average		38.1	17.4	20.9	23.7

6c. Chocolate creams

	Δ Rank	S–S	S–C	C–S	C–C
Pair:					
A–B	0.22	56.2	12.4	18.3	13.0
A–C	1.65	51.5	14.2	19.5	14.8
A–D	2.03	34.3	13.6	23.7	28.4
B–C	1.43	35.5	15.4	23.7	25.4
B–D	1.81	55.6	13.0	17.2	14.2
C–D	0.38	54.4	13.0	13.6	18.9
Hedonic judgments average		47.9	13.6	19.3	19.1

6b. Crackers

Paired comparison	Δ Rank	S–S	S–C	C–S	C–C
Pair:					
A–B	1.84	46.8	14.0	21.1	18.1
A–C	0.65	23.4	25.1	31.0	20.5
A–D	1.08	42.7	19.3	18.1	19.9
B–C	1.19	39.2	18.7	24.6	17.5
B–D	0.76	32.2	22.2	19.9	25.7
C–D	0.43	35.7	22.8	22.2	19.3
Paired comparison average		36.7	20.3	22.8	20.2

6d. Chocolate creams

Paired comparison	Δ Rank	S–S	S–C	C–S	C–C
Pair:					
A–B	0.41	50.9	16.2	18.0	15.0
A–C	1.47	44.3	16.8	18.0	21.0
A–D	1.80	21.0	22.8	22.8	33.5
B–C	1.06	29.9	22.2	20.4	27.5
B–D	1.39	55.7	16.8	16.8	10.8
C–D	0.34	41.9	12.6	22.2	23.4
Paired comparison average		35.6	17.9	19.7	21.9

Crackers: Hedonic judgment (6a) and Paired comparison (6b). Chocolate creams: Hedonic judgment (6c) and Paired comparison (6d). Percentages of children remaining stable in their preference over three sessions (S–S), remaining stable over the first two sessions and then changing (S–C), changing after session 1 and then remaining stable (C–S) and changing after session 1 and changing back after session 2. Also indicated are the differences in preference rank of the two stimuli in a pair (Δ Rank).

As can be seen from these tables the average percentage of children that stick to their first preference (S–S) lies below 50% in all cases and even below 40% in three of them. Furthermore these percentages seem to be somewhat higher with the hedonic categorization method than with paired comparison, especially in the case of the chocolate creams. Changes in preference to the other product occur slightly more frequently after the first session (C–S; crackers 20.9 and 22.8%; chocolate creams 19.3 and 19.7%) than after the second session (S–C; crackers 17.4 and 20.3%; chocolate creams 13.6 and 17.9%). Changes away from the original preference and back (C–C) to it, are about as frequent (crackers 23.7 and 20.2%; chocolate creams 19.1 and 20.9%) as the stable changes away from it (C–S). Further inspection shows that although these trends are on average about the same for all pairs, the stability (S-S) may vary from pair to pair. This one might expect on the basis of the differences in general liking of the pair members. If two pair members are either liked or disliked to almost the same degree, one might expect to find lower stability in the preference than if one pair member was liked much more than the other. In order to check this the differences between the ranks (Δ rank) obtained in liking for the two pair members (see Table 5) have also been given in the tables. For the data on crackers, the percentages for stable liking (S–S) seem indeed related to the differences in liking (rank difference), but for the chocolate creams this is not at all the case. Here the pair (A–D) with the highest difference in liking shows the least stability in both methods and the pair with the lowest difference in liking (A–B) in the hedonic judgments shows the highest stability and the same pair, which is the second highest in liking difference in the paired comparison, shows the one but lowest stability with this method. Thus, it can be concluded that the tendency to change preferences is not merely the result of indecisiveness due to the equality of stimulus attractiveness, but that other and sometimes much stronger factors may also come into play. Product boredom and/or slowly growing product aversion might be good candidates for such factors.

Table 7
Percentages of stability–change patterns in children of different ages

	Crackers				Chocolate creams			
	7a. Hedonic judgments				7c. Hedonic judgments			
Age in years	S–S	S–C	C–S	C–C	S–S	S–C	C–S	C–C
5–6	31.2	13.5	26.6	28.7	39.9	14.2	16.1	30.9
6–7	41.7	13.0	25.5	19.9	40.4	21.7	19.2	18.7
7–8	39.7	18.6	20.1	21.6	41.2	13.2	27.9	17.6
8–9	34.5	26.4	10.9	28.2	61.9	7.9	15.1	15.1
9–10	46.9	18.5	16.7	17.9	52.0	12.6	18.7	16.7
	7b. Paired comparison				7d. Paired comparison			
Age in years	S–S	S–C	C–S	C–C	S–S	S–C	C–S	C–C
5–6	32.7	16.4	22.3	28.6	28.4	23.6	26.4	21.6
6–7	37.7	22.0	22.3	18.0	32.1	20.3	22.4	25.2
7–8	34.7	25.3	22.1	17.9	40.3	22.1	16.8	20.9
8–9	34.8	23.8	22.8	18.6	38.8	14.5	24.3	22.4
9–10	41.0	19.3	25.9	13.8	45.8	17.6	16.7	20.0

Crackers (7a, 7b) and chocolate creams (7c, 7d). Hedonic judgments (7a, 7c) and Paired comparison (7b and 7d) (Legend S–S, S–C, etc see Table 6).

On the other hand, it should be remembered that thus far in this discussion only data of children have been used as illustrations and that young children are renowned for their playfulness and for the instability of their behavior. The question then arises whether in the group of children the stability in their choice grows with growing age. Table 7 provides some insight in this question.

As can be see from these tables, irrespective of the method used, the percentages of children that stick to their preference (S–S) are indeed lowest in the 5–6 year olds and highest in the 9–10 year olds, but the differences are not very dramatic and even with the oldest children the percentages S–S remain well under 50%. At the same time the percentage of children who change their preference twice (C–C) and thus return to their original preference, diminishes with increasing age.

The question whether such variable behavior does only occur in children can partly be answered by the results of a methodological experiment by A.S. Marcelino (2000) in which she happened to measure preference for the visual aspects of cookies with adults. Table 8 gives the results found with hedonic categorization (measured with a nine-point hedonic scale) and paired comparison respectively.

Table 8
Visual aspects of cookies: Hedonic judgment (8a) and Paired comparison (8b). (Legend see Table 2)

8a. Hedonic judgment						8b. Paired comparison					
	Percentage change to:						Percentage change to:				
Originally liked:	1	2	3	4	Average	Originally liked:	1	2	3	4	Average
1	–	57.1	44.4	42.1	49.3	1	–	13.8	30.1	28.6	24.3
2	27.6	–	38.5	45.5	36.4	2	36.4	–	37.0	25.2	33.3
3	31.3	33.3	–	34.6	32.1	3	36.0	22.6	–	17.2	25.7
4	16.1	39.3	29.2	–	27.7	4	35.1	19.2	31.3	–	28.4

Data from Marcelino (2000).

In general, the average amounts of change found in the adults (range hedonic judgment 27.7–49.3%; paired comparison 24.3–33.3%) are somewhat lower than those found for the children with the colored biscuits (range hedonic judgment 34.6–52.7%; paired comparison 27.4–40.4%). Whether these somewhat lower percentages are due to the fact that the participants were adults, to differences between the products or to the fact that only visual aspects were judged by the adults, whereas the children also tasted the products, remains open for discussion.

Comparison of the Tables 8a and 8b shows also that when visual aspects are involved, the adults are more stable in their responses in the paired comparisons than in the monadically presented hedonic judgments. In this respect they behaved exactly like the children who were confronted with differently colored biscuits (see discussion of Tables 2a and 2b).

Just as in the case with the children, upon first seeing the percentages change found for the two methods one might think that this result proves that paired comparison is the more reliable and therefore the better one. It is interesting however to look a little deeper as is illustrated in Table 9. In the experiment all possible pairs were given in both orders of presentation of the stimuli within a pair. Table 9a shows the results of the cases in which the stimulus with the lower number was placed at the left and inspected first and Table 9b shows the results of the cases in which the stimulus with the higher number was placed on the left and inspected first. The numbering of the stimuli was arbitrary.

As can be seen the influence of the presentation order of the two stimuli within a pair on the amount of change found between the two sessions is remarkable. Thus, the percentage change in the pair of products 3 and 1 for those who originally preferred product 3, varies from 15.0 to 57.1% depending on the presentation order. This illustrates one of the weak points of paired comparison. Often inspection of the first sample sets the criteria on which the comparison is going to be based and as a consequence, different presentation orders may lead to different criteria in pairs. If this is true for visual aspects, where looking back and forth is possible, it might be all the more so in cases of smelling and tasting where a temporal order is always imposed.

Returning to the two questions posed before this digression into visual aspects, the results of an experiment by Lévy and Köster (1999) may be helpful in finding the answers. They asked 85 regular drinkers of a given alcoholic beverage to judge three slightly different variations of this drink on three successive days. All samples had the same alcohol content and they were

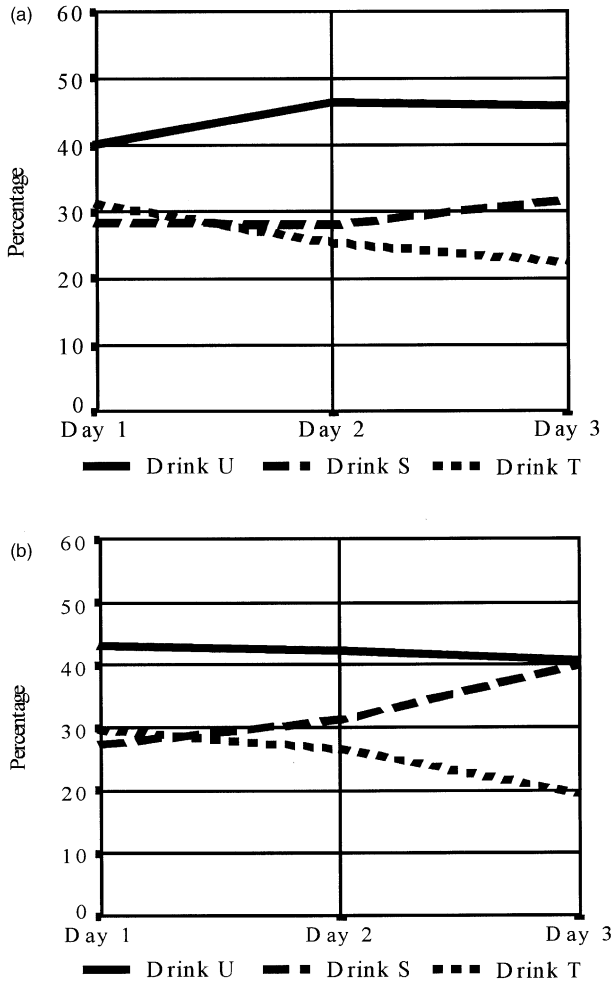

Fig. 1. (a) Percentage of drinks preferred in the hedonic judgment on each of the 3 days of the test ($N=128$). (Observations are weighted according to the number of preferred drinks.) (b) Percentage of drinks finally chosen after consumption on each of the 3 days of the test ($N=128$).

Table 9
Visual aspects of cookies: paired comparison for presentation orders A–B (9a) and B–A (9b). Legend see Table 2

9a. Paired comparison order A–B

	Percentage change to:				
Originally liked:	1	2	3	4	Average
1	–	20.7	25.8	29.6	25.3
2	42.5	–	30.8	21.7	31.9
3	57.1	26.9	–	16.7	32.4
4	44.0	31.0	32.1	–	35.4

9b. Paired comparison order B–A

	Percentage change to:				
Originally liked:	1	2	3	4	Average
1	–	6.9	34.4	27.6	23.3
2	30.4	–	43.3	28.6	34.6
3	15.0	18.2	–	17.6	18.9
4	26.1	8.3	30.4	–	21.4

sent to the home of the participant in nine mini-bottles that each had a different 3-digit code. For the first day, the subjects were instructed to open three bottles with given codes, to pour them in identical glasses and to judge a first sip of them on a nine- point hedonic scale. After this they could drink of them as they liked and finally they were asked to make a choice of the drink they would later want to be sent home to them. Fig. 1a gives the numbers of people who preferred a given drink in the hedonic judgments and in the choice on each of the 3 days of the experiment.

As can be seen from the figures the preference for the drinks changed with both methods over the three days, but the changes were much clearer in the choice at the end of the drinking period than in the hedonic judgments made on the basis of the first sip. In fact drink S, which was somewhat more complex in taste, became liked as much as the more familiar drink U. The preference for drink T steadily deteriorated. Here, it is difficult to say whether the rise in preference for drink S was due to its action as a "pacer" or to the dissipation of neophobia.

In an experiment in which people of East- and West-Germany were exposed to variants of food that were hitherto unknown to them (Köster, Rummel, Kornelson, & Benz, 2001) it could be shown that adult consumers also change their original preference on average in about 50% of the cases, but that the extent to which this happens differs for different types of product.

3. Summary and conclusions

Table 10 gives an overview of the experiments used in this illustration and summarizes the most important outcomes.

Table 10
Overview of the experiments (number of subjects, age, stimuli, methods, number of sessions (#S), results and conclusions)

Experiment	*N*	Age in years	Type of stimuli	Method(s) used	#S	Results and conclusions
Léon et al., 1999 Table 2	157	6– 10	5 Jam topped biscuits 5 tastes (old and new) Different colors	Hedonic Judgments (HJ) (4 point smiley scale) Paired comparison (PC)	2 2	Largest change for novel products. Change PC < HJ. Influence of Color?
Léon, 1998 Table 3	129	6– 10	5 Jam topped biscuits 5 tastes (same as above) All the same red color	Hedonic Judgments (HJ) (4 point smiley scale) Paired comparison (PC)	2 2	Change PC = HJ Influence color and/or familiarity?
Léon, 1998 Table 4	180	6– 10	4 Jam topped biscuits 4 tastes (red fruits, 2 new of which 1 a "pacer") All the same red color	Hedonic Judgments (4 point smiley scale) Paired comparison	2 2	Wild strawberry a "pacer"? Not due to dissipation of neophobia or "mere exposure"
Léon, 1998 Table 5	170 199	6– 10	4 salty crackers 4 chocolate creams	Hedonic Judgments (4 point smiley scale) Paired comparison	3 3	Changes over sessions are cumulative, but not mere sums.
Léon, 1998 Table 6	170 199	6– 10	4 salty crackers 4 chocolate creams	Hedonic Judgments (4 point smiley scale) Paired comparison	3 3	Four types of change behaviour. SS = 39.3%; CS > SC; CC = 21.2%
Léon, 1998 Table 7	170 199	6– 10	4 salty crackers 4 chocolate creams	Hedonic Judgments (4 point smiley scale) Paired comparison	3 3	Stability increases with age but remains < 47%
Marcelino, 2000 Table 8	50	18– 26	4 chocolate cookies (visual inspection only)	Hedonic Judgments (HJ) 9–point hedonic scale Paired comparison (PC)	2 2	Change also in visual inspection by adults. Change PC < HJ (visual)
Marcelino, 2000 Table 9	50	18– 26	4 chocolate cookies (visual inspection only)	Paired comparison (order of presentation compared)	2	Change dependent on presentation order
Lévy and Köster, 1999 Fig. 1a and 1b	128	18–37	3 Alcoholic beverages 3 tastes 1 familiar 2 new (1 simple and 1 complex)	9-point hedonic scale Choice of gift bottle	3 3	Increase of preference for complex drink not due to "mere exposure"

Although more research should be done with adults, they seem to change their preferences after their first judgment in much the same way as the children do. In both groups first preferences seem to be a poor indication of their later choices. This is in agreement with what most of the psychological theories discussed (see Table 1) would predict. The only theory that is not supported is the "mere exposure" theory of Zajonc (1968). Contrary to this theory, repeated exposure may lead not only to an increase, but also to a decrease of the liking for a product. In fact, the "mere exposure" theory can be seen as a special case of the more general theories of Berlyne (1970) and of Dember and Earl (1957), who predict that, when a subject is exposed to a stimulus that is somewhat more arousing (or more complex) than the stimulus he likes best, his optimally liked level of arousal (or complexity) will move in the direction of the level of that stimulus. This and the fact that, as a result of experience, such an arousing or complex stimulus may become a little less arousing (or complex) to the subject, will lead to an increasing liking for stimuli that in the beginning were too arousing or complex to be liked. This would also be predicted by Zajonc. However, according to Dember and Earl, such a change in the optimal level of arousal or complexity of the individual is uni-directional and stimuli that are less arousing or complex than the optimal level can not cause a lowering of the optimal level. At the same time it should be realized, that when exposition to a complex stimuli leads to a shift of the optimum to a higher level, all stimuli below this new optimal level will be liked less than before. Since in the experiments described earlier, the subjects were exposed to all stimuli and since some of these stimuli may have been more complex (e.g. wild strawberry in Table 4) than the initial optimal level, this will have created a shift to a higher level. The liking for other less complex stimuli (e.g. cherry in Table 4) may then have decreased. This decrease would be in accordance with the theory of Dember and Earl (in which wild strawberry would be seen as a "pacer"), but can not be explained by the mere exposure theory of Zajonc.

Another phenomenon that has been discussed already is the difference in amounts of change between the hedonic measurements and paired comparison when visual differences between the stimuli are involved (Tables 2 and 8). Here, it could be argued that the seemingly larger stability of the paired comparison method, is precisely one of its disadvantages, when its validity for most real life situations is concerned. Unless one is standing in front of a counter in a cafeteria, one usually can not see different foods at the same time. This combined with the fact that the order of presentation of the two samples has such a marked influence on the changes in preference (Table 9), seems to plead against the use of the paired comparison method as a predictor for real food choice.

Finally, if individual change is indeed as frequent as might be suspected on the basis of the psychological theories mentioned in the introduction, what does this mean for the concept of repeatability of a method and how can it be measured?

In principle, one may not expect that the same population will say the same thing when offered the same set of samples on two separate occasions. Good repeatability would rather mean that if two similar samples drawn from the same population were tested twice they would change in the same way. A split half measurement comparing the development of the answers of the two parts of a population on both their equality of response in each of the sessions and their equality in the extent and the direction of change over sessions might probably be a solution to this problem.

The details of such a procedure involving multiple split halves and based on individual results need to be worked out. It is likely to provide a better measure of repeatability than the simple reliability coefficients that are still used too often.

References

Berlyne, D. E. (1970). Novelty, complexity and hedonic value. *Perception and Psychophysics*, *8*, 279–286.

Brickman, P., & D'Amato, B. (1975). Exposure effects in a free-choice situation. *Journal of Personality and Social Psychology*, *32*, 415–420.

Cabanac, M. (1971). The physiological role of pleasure. *Science*, *173*, 1103–1107.

Dember, W. N., & Earl, R. W. (1957). Analysis of exploratory, manipulatory and curiosity behaviors. *Psychological Review*, *64*, 91–96.

Dember, W. N., Earl, R. W., & Paradise, N. (1957). Response by rats to differential stimulus complexity. *Journal of Comparative Psychology*, *49*, 93–95.

Dember, W. N. (1970). *Stimulus complexity, motivation and emotion. The psychology of perception* (2nd ed. pp. 341–375). London: Holt, Richart and Winston.

Engen, T. (1971). Psychophysics. In J. W. Kling, & L. A. Riggs (Eds.), *Experimental psychology*. New York: Holt Rinehart and Winston.

Glanzer, M. (1953). Stimulus satiation: an explanation of spontaneous alternation and related phenomena. *Psychological Review*, *60*, 257–268.

Harrisson, A. A., Tutone, R. M., & McFadgen, D. G. (1971). Effects of frequency exposure of changing and unchanging stimulus pairs on affective ratings. *Journal of Personality and Social Psychology*, *20*(1), 102–111.

Köster, E. P. (1990). Nutrition psychology. In *Food ingredients Europe Conference Proceedings 1990* (pp. 32–37). Maarssen, The Netherlands: Expoconsult Publishers.

Köster, E. P. (1991). The dynamics of consumer preferences and aversions. In *Food ingredients Asia Conference Proceedings 1991* (pp. 223–227). Maarssen, The Netherlands: Expoconsult Publishers.

Köster, E. P., Rummel, C., Kornelson, C., & Benz, K. H. (2001). Stability and change in food liking: food preferences in the two Germany's after the reunification. In M. Rothe (Ed.), *Flavour 2000: perception, release, evaluation, formation, acceptance, nutrition and health*. Germany: Rhothe, Bergholz-Rehbrücke.

Lähteenmäki, L., & Van Trijp, H. C. N. (1995). Hedonic responses, variety-seeking tendency and expressed variety in sandwich choices. *Appetite*, *24*, 139–152.

Léon, F. (1998). *Elaboration et comparaison de méthodes d'évaluaton des préférences alimentaires adaptées aux enfants de quatre à dix ans*.

(Establishment and comparison of methods for the evaluation of food preferences in children from four to ten years). Doctoral thesis, Université Paris 6, France.

Léon, F., Couronne, T., Marcuz, M. C., & Köster, E. P. (1999). Measuring food liking in children: a comparison of non verbal methods. *Food Quality and Preference*, *10*, 93–100.

Lévy, C., & Köster, E. P. (1999). The relevance of initial hedonic judgements in the prediction of subtle food choices. *Food Quality and Preference*, *10*, 185–200.

Marcelino, A. S. (2000). *Contribution a l'étude du caractère appétissant des aliments: role des facteurs visuels. (Contribution to the study of the appetizing character of food: the role of visual factors*). Doctoral thesis, ENSIA, Massy, France.

Pliner, P. (1982). The effects of mere exposure on liking for edible substance. *Appetite*, *3*, 283–290.

Punter, P. H. (1976). Odor intensity for *n*-butanol: individual versus group data. *Proceedings Second ECRO Congress, Reading, UK* (abstract).

Punter, P. H., & Köster, E. P. (1977). The intensity slopes of the n-aliphatic alcohols. In J. Le Magnen, & P. MacLeod (Eds.), *Olfaction and taste VI* (pp. 71). London: IRL.

Richman, C. L., Dember, W. N., & Kim, P. (1986). Spontaneous alternation behavior in animals: a review. *Current Psychological Research and Reviews*, *5*–1987, 358–391.

Rolls, B. J., Rowe, E. A., Rolls, E. T., Kingston, B., Megson, A., & Gunary, R. (1981). Variety in a meal enhances food intake in man. *Physiology and Behavior*, *26*, 215–221.

Tolman, E. C. (1925). Purpose and cognition: the determinants of animal learning. *Psychological Review*, *32*, 285–297.

Van Trijp, H. C. M. (1984). Product related determinants of variety seeking behavior for foods. *Appetite*, *22*, 1–10.

Young, P. T. (1968). Evaluation and preference in behavioral development. *Psychological Review*, *75*, 222–241.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of personality and social psychology*, *9*(Monograph Suppl.), 65–74.