Eine Zeitschrift aus dem Gabler Verlag • 70610

GABLER
2|2007

Thexis

Fachzeitschrift für Marketing

Universität St. Gallen

Produktdesign

02

Design Evaluation:
From Typical Problems to State-of-the-Art Solutions

Most consumer products have to fulfil a variety of requirements to appeal to consumers and ensure market success. Today, one key factor is the design of the product. Especially for products in highly competitive markets, products with high costs of development and long product renewal intervals, new design concepts have to be chosen very deliberately. There is a thin red line between being too innovative and being too conservative, the former causing reluctant consumer behaviour and the latter generating boring and disliked products in the long run. Therefore, sophisticated and everyday-life relevant techniques of design evaluation have to be used. The present article describes typical problems of standard evaluation strategies and develops solutions to overcome them. The *repeated evaluation technique* is proposed as a state-of-the-art tool not only for measuring current preferences but also to assist with predicting future appreciation of consumer designs.

Assoc. Prof. Dr. Claus-Christian Carbon
Associate Professor at the Department of Psychological Basic Research, Faculty of Psychology at the University of Vienna, Wien (A)

Prof. Dr. Helmut Leder
Head of the Institute of Psychological Basic Research, Faculty of Psychology at the University of Vienna, Wien (A)

Importance of aesthetic design aspects

In competitive markets that offer products with highly interchangeable technical aspects (such as cars, hi-fi components, household appliances, cellphones, notebooks, etc.), product design is an essential factor for market success (Kreuzbauer/Malter 2005). Appropriate product designs require a tight fit between consumers' interests, wishes and desires and the design. Concerning functional and technical design aspects, a great variety of tests on usability (Jordan 1998), human factors (Green/Jordan 1998) and ergonomics (Salvendy 2006) have been developed over the last 30 years. These are extensively used by consumer product manufacturers. However, for aesthetic and pleasure-based aspects of product designs neither standard methods have been developed nor are they used in a standardized or systematic way (Hekkert 2006; Jordan 2000). In many cases, aesthetic dimensions are ignored altogether (Liu 2003). The usability expert Patrick W. Jordan terms such aesthetic dimensions amongst "New Human Factors", which strikingly demonstrates the importance and the newness of this approach. Future design evaluations will not be complete without taking aesthetic design aspects into account (cf. the contribution by Leder, Carbon and Kreuzbauer in this issue of Thexis)!

Focus on aesthetics in design

A remarkable success story of pronouncing aesthetic aspects for a consumer product of highly interchangeable technical aspects is the Apple iPod. Since the very early beginnings of iPod in October 2001, the design followed a strict product philosophy of being aesthetical, innovative and stylish. From a pure technical or monetary standpoint, many competing portable media players have higher functionality, more flexible user interfaces and a better cost effectiveness. However, presumably due to iPod's high aesthetic and stylish properties, it is currently the world's best-selling digital audio player and makes it to one of the most popular consumer brands. Consequently, the iPod today is an essential economic backbone of Apple Computer Inc.

Problems of measuring aesthetic appreciation

An increasing number of companies are beginning to focus on aesthetic design aspects, however, they often do so without standardized assistance on how to measure aesthetic qualities. Typically, products to be developed, concept products, prototypes or brand new products are evaluated in extensive and costly tests with typical consumers in different test settings: *car clinics* in the automobile sector, or focus groups, think-aloud protocols and questionnaires (see Jordan 2000). It is a general problem that all these tests measure consumers' evaluations only *once* in "single-shot tests". Moreover, the material mostly is quite unfamiliar to the consumer. This type of evaluation technique leads to biased responses which do not necessarily reflect the typical everyday experiences with product designs. For example, Leder

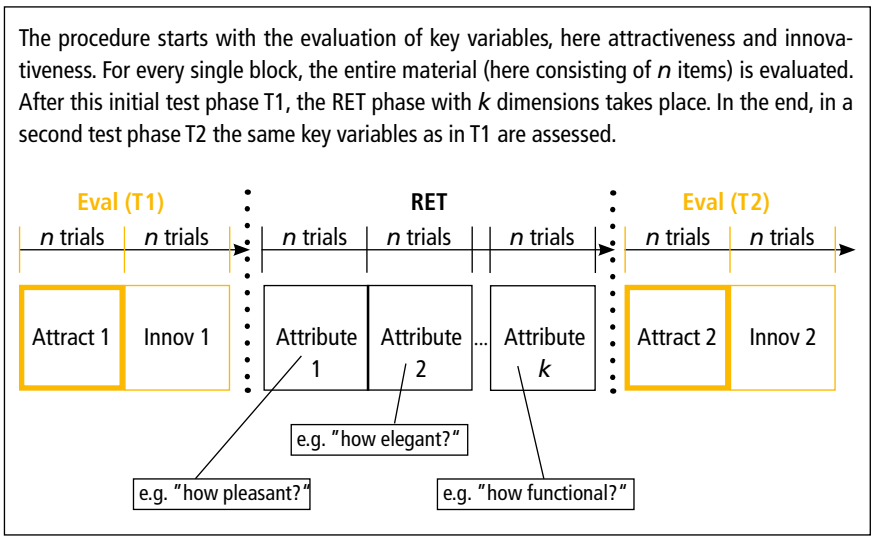

Fig. 1: Schematic time course of a RET procedure for evaluating consumer products
Source: Carbon/Leder 2005.

and Carbon (2005) using material varying in terms of innovativeness have recently shown that participants being unfamiliar with the material strongly prefer familiar material which is quite conservative, while rejecting material of higher levels of innovativeness. However, successful products, such as the Apple iPod, or recent developments in the automobile sector, are above all, highly innovative in their designs.

Simulation of everyday life experiences

How can it be explained that people prefer quite conservative material in experimental studies but often tend to prefer innovative material in everyday life? Carbon and Leder (2005) have proposed that everyday life experiences first have to be simulated *before* one can measure preferences or liking on a valid basis. If this principle is overlooked, invalid and quite misleading predictions for future preferences are made. As a consequence, future developments of consumer products assimilate to these invalid or at least very limited evaluations, which are based on first-glance responses of consumers who had no chance to familiarize themselves with these products. This can turn out to be fatal, as such product developments do not consider the dynamics of everyday life experiences. Consumer products which have been developed on the basis of such inadequate pre-evaluations therefore have a high risk of a short market life, not penetrating the markets in the long run and, in the end, even becoming economic failures.

Obviously there is substantial need for a more ecologically valid measurement technique that *firstly* simulates everyday life by letting typical consumers elaborate the material and *secondly* allows consumers to evaluate the material after familiarization.

The repeated evaluation technique (RET)

Carbon and Leder (2005) have developed the *repeated evaluation technique (RET)* which integrates both requirements, (1) deep elaborations and (2) valid evaluations of the material, into one single procedure. The integral procedure of *RET* is illustrated in Figure 1. It consists of two identical test phases (T1 and T2), where key variables are measured (e.g., attractiveness, innovativeness, liking, etc.), and an intermediate elaboration phase, where repeated evaluations of the entire material are requested from the participants.

By employing the *RET* with car interior designs that varied on different levels of curvature, complexity and innovativeness, they demonstrated that typical consumers are only initially rejecting highly innovative material while preferring familiar or conservative material (Leder/Carbon 2005; Zajonc/Markus 1982). After having elaborated the material via *repeated evaluations* and *elaborations*, they tend to prefer innovative material more and reject more conservative material. Thus, conservative material presumably turns out to be boring after a while. This is in accordance with the influential theory of D. E. Berlyne who proposed that interest, novelty and curiosity are important predictors for exploratory behaviour that in turn directly affects preferences (Berlyne 1970), as well as with recent evidence from market research (Kinnear/de Kock 2006; Zandstra et al. 2004).

These elaboration-dependent effects are illustrated in Figure 2: innovative designs might be misleadingly evaluated in simple single-shot studies as being relatively unattractive and conservative designs, in contrast, as being relatively attractive (illustrated as T1). However, by employing the *RET*

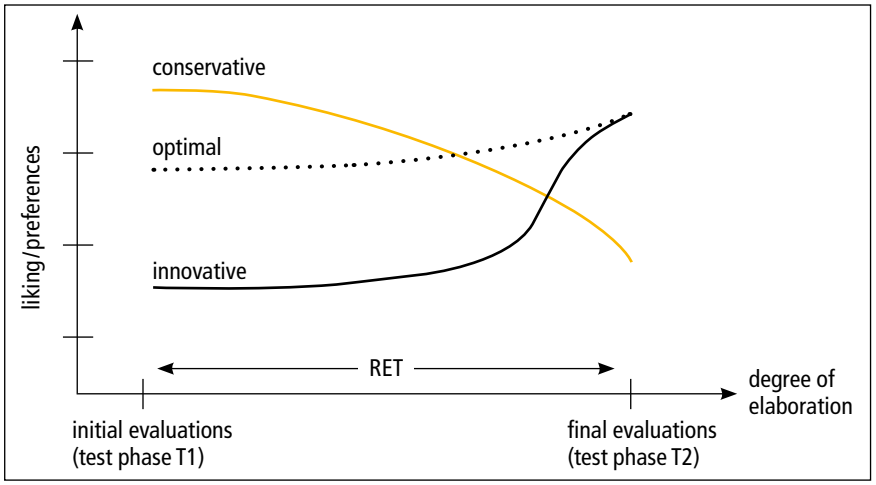

Fig. 2: Idealized dynamic development of different qualities of consumer product designs empirically supported by the RET
Source: Carbon/Leder 2005.

On the left side, you can see a typical inverted U-shaped curve which can be obtained at evaluation phase T1: people tend to reject unfamiliar while favouring conservative material; on the right side, you can see a butterfly curve which can be obtained at evaluation phase T2: after having elaborated the material, people do not prefer the conservative (familiar) material anymore but choose more innovative material. Optimal material is indicated by –U(nfamiliar) and +U(nfamiliar). For reasons of comparison, these two points are also shown in the left diagram.

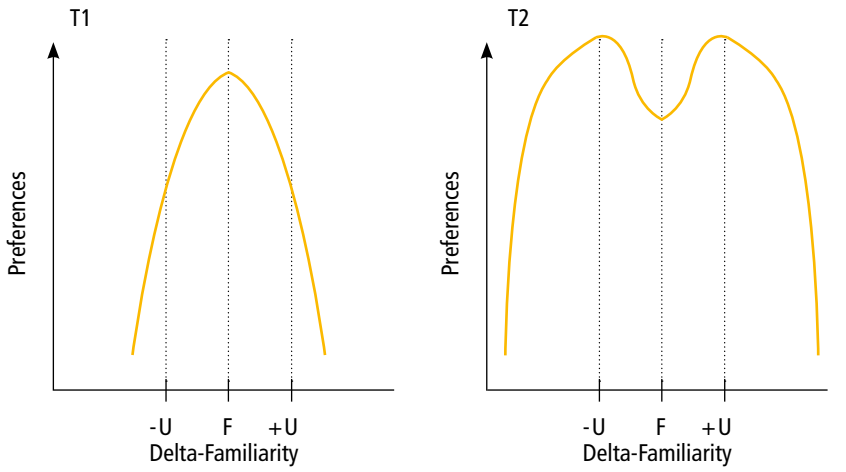

Fig. 3: Dynamics of preferences for material varying in terms of delta-familiarity (difference of familiarity from the product to be evaluated and the perfectly familiar reference product F)

a very different response pattern might be found in T2. Measuring preferences at T2, after having elaborated and understood the material (the *RET* block), might turn the relations between innovative and conservative material at T1 upside down. Now the innovative material is liked much more, while the conservative design looses its appeal. As a third curve, an "optimal" progression over elaboration is portrayed. Optimal material progresses over time as shown by the dotted line. It is characterized by an optimal combination of both familiarity[1] and innovativeness (cf. Hekkert et al. 2003). Thus, having equilibrium between both dimensions, such material is rather positively evaluated when seen initially, and gains attractiveness with increasing elaboration.

What we can learn from RET

As pointed out above, typical single-shot studies in which preferences are measured only once, *before* participants have elaborated materials of interest, are limited to first-glance assessments and cannot predict future preferences (Carbon 2005, 2006). How preferences tend to change dynamically over time is illustrated in Figure 3 where the preferences for objects, differing to the degree of familiarity, are shown in dependence of the difference of familiarity to a perfectly familiar object which is liked (the *delta-familiarity*).

When the material is evaluated for preferences in T1, there is a clear-shaped inverted U-curve: every object deviating from the perfectly familiar object is more or less disliked. However, after employing the *RET*, a butterfly-shaped curve will emerge with maxima left and right of the perfectly familiar object. Interestingly, after having elaborated the material, these maxima can even outrange the maximum of the perfectly familiar object in T1. Such a description of changes in preferences, due to dynamic aspects of elaboration, points to the necessity of applying methods that capture dynamical changes.

A typical test scenario with RET

In the following, we will portray a typical design evaluation study based on the *RET*. In the portrayed fictive study, the preferences of different steering wheels are the focus of research[2]. It is assumed that the material is varied on two basic dimensions: curvature and innovativeness, both dimensions on three levels: low, medium and high. As illustrated in Figure 4, both dimensions are fully combined in all tested steering wheels, resulting in 3 x 3 = 9 objects, all of which have to be evaluated. The key variable of interest will be the preference for the consumer products. According to the *RET* approach, the preferences will be measured in T1, right before, and in T2, right after the *RET*, for every single steering wheel. In the intermediate *RET* phase, illustrated in Figure 4, a variety of further attributes, which should help to elaborate the material, have to be evaluated. To ensure elaboration, these attributes should be pre-selected by experts on the chosen topic. Here, attributes such as "how pleasant", "how elegant" or "how functional" could be used. Typically, at least 25 attributes are recommended (cf. Carbon et al. 2006; Carbon/Leder 2005). The number of participants depends on material, the relevant populations and the statistical parameters relevant for the effect size (see also Erdfelder et al. 1996). As participants, typical and potential consumers of a target product are preferable. It is important to stress that *RET* does neither need specific levels of expertise towards a target product nor specific levels of adoption behaviour (in the sense of Ryan/Gross 1943).

Results obtained would range from typical "at first glance" evaluations from T1 on the one hand to more ecologically valid evaluations from T2 on the other hand. The evaluations of T2 can be interpreted as predictions of object preferences in the nearer future.

Extensions of RET

The *RET* can be used in classical behavioural contexts, such as evaluation or questionnaire studies, but has also proved appropriate when combined with more sophisticated techniques of market research.

[1] In RET studies the factor complexity, which was identified as an important dynamical factor for preferences over time (for 'stimulus complexity' see Berlyne 1970; for 'stimulus complexity' and 'individual's complexity' see Dember/Earl 1957), should be controlled for the material.

[2] RET is principally not limited to a certain class of stimuli. Nevertheless, up to now, it was only tested for visually presented stimuli, but not for auditory, tactile or olfactory material.

Typical materials for using RET in market research, here with steering wheels varying on two base dimensions: curvature and innovativeness. The third dimension shows the different attributes asked in the RET procedure.

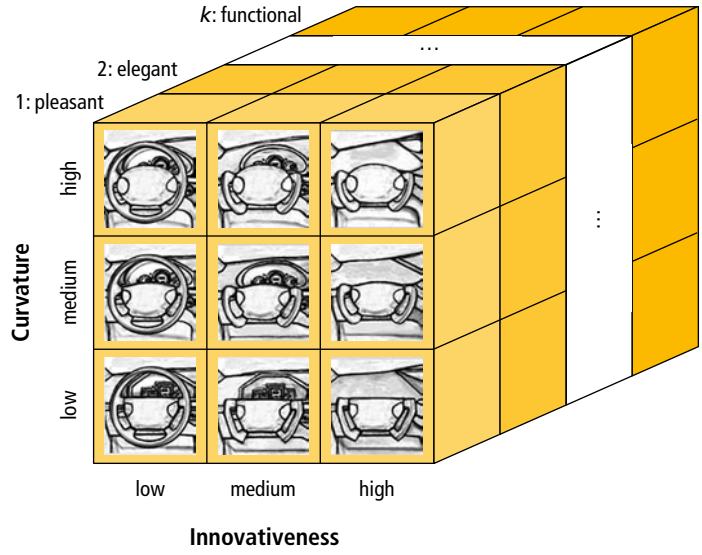

Fig. 4: Typical materials for using RET in market research

Pupillometry

For example, Carbon et al. (2006) have recently shown that *RET* can easily be used with eye-tracking devices measuring pupillometry, the size of the pupils. Pupillometry provides an interesting opportunity to analyze affective states, because the diameter of the pupil cannot be regulated cognitively. Thus, pupillometry is not confounded by conscious components. Carbon et al. (2006) revealed that beside behavioural indications of preferences for innovative designs, there is also support for dynamically changing preferences in pupillometry data: there was a greater increase of the pupil diameter for innovative than for conservative material, but this was only seen in T2!

Eye-tracking

Moreover, analysis of eye-movements showed that participants followed more harmonic scanpaths when being exposed to innovative material (Carbon et al. 2006). Carbon et al. speculated that the higher *visual rightness* (Locher 2003), the inner quality of a picture in terms of balance, only becomes obvious after elaboration of the material. Thus, this quality becomes evident when tested after the *RET*.

Skin conductance

Parallel findings stem from other work in the area of psychophysiology. Carbon and colleagues have used *electro-dermal activity* (EDA), which is a measure of skin conductance response, in combination with *RET*. The EDA is highly sensitive to emotional activity, attention processes and stimulus significance (Dawson et al. 2000). Technically speaking, EDA measures the eccrine sweat glands which are neurally entirely under sympathetic control and which are processed by early automatic discrimination processes (Lyytinen et al. 1992). Thus, EDA is similarly as pupillometry cognitively not penetrable, which makes it a highly valuable tool for measuring non-conscious preferences. Measuring EDA within the *RET* paradigm revealed an increase of electro-dermal activity only for those materials which are preferred in test phase 2, right *after* the *RET* ratings once again, participants showed only specific, thus, indicative responses after having deeply elaborated the material.

Conclusions and outlook

To summarize, we have developed a method *(RET)* which is a valuable method to understand dynamic changes in the aesthetic appreciation of deeply elaborate material. This is the pre-condition for testing preferences of consumer products on a valid basis. Preferences measured by *RET* do not only contain first-glance responses but also responses on the basis of deep elaboration, which are more suitable for predicting future preferences. For proper market research about the quality of products, *RET* therefore seems to be a highly relevant technique.

This research was supported by a grant to HL and CCC from the FWF "Fonds zur Förderung der wissenschaftlichen Forschung" (National Austrian Scientific Fonds; P18910). We thank Andrea Lyman for proofreading this manuscript.

References

Berlyne, D. E. (1970): Novelty, complexity, and hedonic value, in: Perception and Psychophysics, Vol. 8, No. 5-A, pp. 279-286.

Carbon, C. C. (2005): Innovation in Design and Aesthetics. How attributes of Innovation influence Attractiveness on the long run, in: Perception, Vol. 35 (S), p. 8.

Carbon, C. C. (2006): The Repeated Evaluation Technique or ‚How can we measure attractiveness in a valid way', in: Perception, Vol. 35 (S).

Carbon, C. C./Hutzler, F./Minge, M. (2006): Innovation in design investigated by eye movements and pupillometry, in: Psychology Science, Vol. 48, No. 2, pp. 173-186.

Carbon, C. C./Leder, H. (2005): The Repeated Evaluation Technique *(RET)*. A method to capture dynamic effects of innovativeness and attractiveness, in: Applied Cognitive Psychology, Vol. 19, No. 5, pp. 587-601.

Dawson, M. E./Schell, A. M./Filion, D. L. (2000): The electrodermal system, in: Cacioppo, J. T./Tassinary, L. G./Berntson, G. G. (Eds.): Handbook of Psychophysiology , 2nd ed., Cambridge/MA.

Dember, W. N./Earl, R. W. (1957): Analysis of Exploratory, Manipulatory, and Curiosity Behaviors, in: Psychological

Review, Vol. 64, No. 2, pp. 91-96.

Erdfelder, E./Faul, F./Buchner, A. (1996): G*Power: a general power analysis program, in: Behavior Research Methods, Instruments, & Computers, Vol. 28, No. 1, pp. 1-11.

Green, W. S./Jordan, P. W. (1998): Human Factors in Product Design: Current Practice and Future Trends, London.

Hekkert, P. (2006): Design aesthetics: principles of pleasure in design, in: Psychology Science, Vol. 48, No. 2, pp. 157-172.

Hekkert, P./Snelders, D./van Wieringen, P. C. W. (2003): ‚Most advanced, yet acceptable': Typicality and novelty as joint predictors of aesthetic preference in industrial design, in: British Journal of Psychology, Vol. 94, No. 1, pp. 111-124.

Jordan, P. W. (1998): An Introduction to Usability. London.

Jordan, P. W. (2000): Designing Pleasurable Products: An Introduction to the New Human Factors. New York.

Kinnear, M./de Kock, H. L. (2006): Effect of repeated exposure on consumer preferences for sports drinks containing different acidulants: European Sensory Network.

Kreuzbauer, R./Malter, A. J. (2005): Embodied cognition and new product design: Changing product form to influence brand categorization, in: Journal of Product Innovation Management, Vol. 22, No. 2, pp. 165-176.

Leder, H./Carbon, C. C. (2005): Dimensions in appreciation of car interior design, in: Applied Cognitive Psychology, Vol. 19, No. 5, pp. 603-618.

Liu, Y. (2003): Engineering aesthetics and aesthetic ergonomics: theoretical foundations and a dual-process research methodology, in: Ergonomics, Vol. 46, No. 13-14, pp. 1273-1292.

Locher, P. J. (2003): An empirical investigation of the visual rightness theory of picture perception, in: Acta Psychologica, Vol. 114, No. 2, pp. 147-164.

Lyytinen, H./Blomberg, A. P./Näätänen, R. (1992): Event-related potentials and autonomic responses to a change in unattended auditory stimuli, in: Psychophysiology, Vol. 29, No. 5, pp. 523-534.

Ryan, B./Gross, N. C. (1943): The diffusion of hybrid seed corn in two Iowa communities, in: Rural Sociology, Vol. 8, No., pp. 15-24.

Salvendy, G. (2006): Handbook of Human Factors and Ergonomics, 3rd ed., Hoboken/NJ.

Zajonc, R. B./Markus, H. (1982): Affective and cognitive factors in preferences, in: Journal of Consumer Research, Vol. 9, No. 2, pp. 123-131.

Zandstra, E. H./Weegels, M. F./Van Spronsen, A. A./Klerk, M. (2004): Scoring or boring? Predicting boredom through repeated in-home consumption, in: Food Quality and Preference, Vol. 15, No. 6, pp. 549-557.