Applied Ergonomics 39 (2008) 71–85

APPLIED ERGONOMICS

www.elsevier.com/locate/apergo

Designing interactive consumer products: Utility of paper prototypes and effectiveness of enhanced control labelling

Juergen Sauer[a,*], Holger Franke[b], Bruno Ruettinger[b]

[a]*Department of Psychology, University of Fribourg, Rue de Faucigny 2, CH-1700 Fribourg, Switzerland*
[b]*Institute of Psychology, Darmstadt University of Technology, Alexanderstrasse 10, D-64283 Darmstadt, Germany*

Received 6 February 2006; accepted 5 March 2007

Abstract

The studies reported here examined differences in user behaviour when presented with a low-fidelity paper prototype compared to fully operational product, and evaluated the effectiveness of different types of enhanced labelling of controls. In the first study with a paper prototype, 30 users of high-pressure washers were asked to choose the settings of the temperature control for different cleaning objects, comparing standard with information-enriched control labelling. In the second study, 34 users operated a real high-pressure washer with different forms of control labelling. The results of both studies provided evidence for some benefits of an information-enriched control labelling over traditional temperature-centred controls labelling. Furthermore, a comparative analysis of the data of the two studies suggested that low-fidelity paper prototypes may have to be used with caution. Therefore, designers need to be aware that the behavioural effects induced by different design modifications may be overestimated when using paper prototypes. The implications of the findings are discussed within the framework of an enlarged concept of fidelity.

Keywords: Paper prototype; Simulation fidelity; Consumer product

1. Introduction

In the context of consumer product design, a major goal has been to provide human factors expertise as early as possible in the design process (see Chapanis, 1996). This is because the earlier design changes are made, the less costly they are. Therefore, there has been a great interest in the development and use of human factors methods that may be applicable well before a fully operational prototype of the product is available. These methods can be solely expert-based (e.g., heuristic evaluation, checklists) or they may involve the collection of data from users (e.g., paper prototype, computer simulation, mock-up). Although expert-based methods provide some helpful guidance in the design of systems, there are generally stronger benefits to be gained from employing methods that involve current or future systems users. ''User-based'' methods require the availability of some model of the system if the real system is not yet available. The degree to which a model of the system resembles the target system refers to the *fidelity* of the model. The fidelity of the model (or *prototype fidelity*) may vary considerably, ranging from a low-fidelity simulation of the system (e.g., paper prototype) to a fully operational prototype, which is (almost) identical to the real system.

1.1. Prototype fidelity

The use of low-fidelity prototypes in industry is widespread because they offer a number of advantages, such as rapid availability and low development cost (Rudd et al., 1996). At the same time, there are concerns that low-fidelity prototypes are only of limited utility for the identification of usability problems because they may invoke different user reactions and behaviour than the actual product. It may also be of importance to what aspect of the consumer product the level of prototype fidelity refers. According to a classification system suggested by Virzi et al. (1996), one may distinguish between four dimensions of fidelity: degree

*Corresponding author. Tel.: +41 26 3007622; fax: +41 26 3009712.
E-mail address: juergen.sauer@unifr.ch (J. Sauer).

0003-6870/$ - see front matter
doi:10.1016/j.apergo.2007.03.001

of functionality, similarity of interaction, breadth of features, and aesthetic refinement. Another classification system distinguishes between four other dimensions of fidelity: physical, functional, cognitive and construct (Elliot et al., 2004). However, to our knowledge, there is no empirical work that examined the respective importance of these dimensions outlined by the different classification systems for prototype evaluation. More generally, even though paper prototypes are widely used in industry, there is a great paucity of rigorously conducted empirical research that examined the pros and cons of different kinds of prototypes. A literature search revealed 10 published studies, which are summarised in Table 1.

A review of the studies revealed that very different aspects of the issues surrounding prototype fidelity were examined. Overall, the studies showed great variations in the kind of prototype (e.g., paper, cardboard, computer simulation) and type of target product used (2D, 3D), resulting in a number of different combinations (e.g., 2D prototype modelling 2D product, 3D prototype modelling 3D product). It is also noteworthy that most of the studies focused on the identification of usability errors while none of them actually examined efficiency measures (e.g., task completion time, energy consumption). The reason for this may be that paper prototypes are generally not considered to be suitable for collecting efficiency measures (Jordan, 1998). Most studies (I, III–VIII) concluded that the prototypes used showed equivalent results whereas studies II, IX and X found some advantages for medium-fidelity prototypes over low-fidelity prototypes.

The review of the studies in Table 1 raises two questions. Most studies were concerned with usability errors rather than efficiency measures. This raises the question to what extent would low-fidelity paper prototypes be suitable to predict user behaviour with regard to efficiency criteria. Furthermore, the review showed that most 2D prototypes had modelled 2D products rather 3D products. This raises the question to what extent would a 2D paper prototype be suitable to predict user behaviour for a fully operational 3D product.

1.2. Control labelling

Since consumer products are typically employed by users who have not undergone any formal training in how to use them, a major concern of system designers has been how critical information can be best conveyed to the user. There are several possibilities available to the system designer, such as instruction manuals, on-product information, hotlines and retail salespeople. Among these, the most widely used means to convey information to the user are instruction manuals and on-product information. Whereas instruction manuals allow the transmission of very detailed information to the user, their major drawback is that they are frequently not read (Sanders and McCormick, 1993). Therefore, critical user information is better conveyed by means of on-product information since it is permanently visible to the user. This may result in higher compliance rates for on-product information than instruction manuals. While on-product information typically refers to labels attached to the product providing users with different forms of advice or warnings (e.g., McCarthy et al., 1995), it may also be integrated into displays and controls. Integrating this information into the controls has the advantage of making a stronger link between the behavioural advice given and the control device needed to implement the advice. On-product information is considered to be effective if it successfully passes the following four information-processing stages (Wogalter, 1999): users do not only need to (a) notice and (b) understand the information, it also needs to (c) correspond to the users' beliefs and attitudes and (d) must motivate users to comply with the message. A number of studies (summarised in a meta-analysis by Cox et al., 1997) demonstrated the overall effectiveness of on-product warnings on a range of technical products (e.g., car, computer disc drive, circular saw, computer software). Outside the context of product warnings, on-product information has been successfully used to encourage environmentally friendly user behaviour for vacuum cleaner and kettle operation (Wiese et al., 2004; Sauer et al., 2003).

Many interactive consumer products with continuous power controls do not provide information about appropriate control settings since their labelling is often limited to quantitative information in the form of physical units (e.g., temperature in °C). This may be helpful for expert users but novice users may be better supported by qualitative labels providing advice about appropriate control settings. An example may be the labelling of the power controls of a vacuum cleaner providing information on power setting (in W) as well as on the objects to be cleaned (e.g., carpet, curtain). In the absence of such qualitative labelling, novice users are likely to use maximum settings, as they often falsely believe that maximum power equals maximum effectiveness (Sauer et al., 2004).

1.3. The present studies

Against this background, the first goal of the present work was concerned with the question to what extent user behaviour with a 3D consumer product can be predicted from user behaviour with a 2D paper prototype. This represents an important methodological issue because it determines the suitability of 2D paper prototypes as a simulation tool. The second goal pursued in the present study was concerned with the optimisation of product design. For that purpose the effectiveness of different design options for control labelling was evaluated to provide better support to users during operation of interactive consumer products.

These two goals are pursued by examining the impact of prototype fidelity and control labels on a range of dependent variables, with a particular focus being placed

Table 1
Comparative studies that examined usability of prototypes of different fidelity levels

Study	Product examined	Fidelity of prototype			Main outcome measures	Findings of study	Reference
		Low[a]	Medium[b]	High[c]			
I	Calendar system, touch screen ticket machine (both 2D)	× (2D)	× (2D)	—	Usability errors	• No difference between paper prototype and computer-based simulation	Sefelin et al. (2003)
II	Videotext system (2D)	× (2D)	× (2D)	—	Usability errors	• No difference in the overall number of usability problems identified but computer simulation identified more of the major usability problems	Nielsen (1990)
III	Cooker controls (3D)	× (2D)	× (2D)	—	Usability errors	• Paper prototype and computer-based simulation identified the same controls arrangement as the most favoured option • Rank order of three alternative controls arrangements differed	Hsu and Peng (1993)
IV	Electronic book, voice response system (both 2D)	× (2D)	—	× (2D)	Usability errors	• No difference between paper prototype and actual product	Virzi et al. (1996)
V	Library search computer system (2D)	× (2D)	—	× (2D)	Usability errors and user satisfaction	• No differences as a function of prototype fidelity (number and severity of usability problems, subjective evaluation)	Catani and Biers (1998)
VI	Electronic dictionary (2D)	× (2D)	—	× (2D)	Usability errors and user satisfaction	• Little difference as a function of prototype fidelity (usability errors, aesthetics)	Wiklund et al. (1992)
VII	On-line banking website (2D)	× (2D)	—	× (2D)	Usability errors	• No difference in number of usability errors between paper prototype and computer-based simulation but users made more comments in the computer simulation condition	Walker et al. (2002)

Table 1 (*continued*)

Study	Product examined	Fidelity of prototype			Main outcome measures	Findings of study	Reference
		Low[a]	Medium[b]	High[c]			
VIII	Drink can refund machine (3D)	× (3D)	—	× (3D)	Usability errors	• No difference between 3D paper prototype and actual product	Säde et al. (1998)
IX	Domestic lighting controller (3D)	× (3D)	× (2D)	—	Usability errors	• The interactive 2D touch screen prototype revealed more usability problems than the 3D cardboard mock-up	Hall (1999)
X	Blood pressure monitor (3D)	× (2D)	× (3D)	× (3D)	Usability errors	• Usability errors identified with real product were more similar to usability errors identified with paper prototype than with 3D mock-ups fidelity • More frequent interaction with real product than with two prototypes of reduced fidelity	Rooden (1999)

[a]Paper or cardboard prototype.
[b]Computer-based simulation or 3D mock-up.
[c]fully operational prototype or commissioned product; 2D, two-dimensional; 3D, three-dimensional.

on efficiency measures. Efficiency measures are typically taken in late phases of the product development process and usually require a high fidelity prototype which is almost equivalent to a fully operational appliance (Jordan, 1998). However, some efficiency measures can also be taken when using low-fidelity paper prototypes (e.g., settings of temperature controls, mowing height of lawn mower blade). Both efficiency measures may also represent indicators of environmental friendliness of user behaviour (e.g., water consumption, energy usage). These were considered to be particularly important because environmentally friendly product usage represents a growing concern in the design of consumer products due to their increasing proliferation (Wenzel et al., 1997). For most consumer products, the most important determinants of environmental friendliness of interactive products relate to the consumption of energy and water as critical resources.

For the present work, the high-pressure washer was chosen as a model product. The high-pressure washer was considered to be suitable for this purpose because of its increasingly widespread use as a cleaning tool, which is characterised by considerable resource consumption during operation (i.e. water, electricity and, in some cases, petrol). For example, nominal energy consumption can reach up to 6.8 kWh for some models and water consumption can be at a rate of up to 1000 L/h during operation.

In order to address the two research goals of the present work (i.e. evaluation of utility of 2D paper prototypes for 3D products and providing design recommendations for control labelling), two experimental studies were carried out, preceded by a pilot survey. The pilot survey was conducted to analyse the general user requirements for high-pressure washer usage. Study I examined different design options, employing a 2D paper prototype. Study II looked at similar design options but used a fully operational product. Finally, a comparative analysis on the merged data sets of the two studies is carried out to evaluate the utility of paper prototypes for modelling 3D products.

2. Pilot study

The purpose of the pilot study was to gain an understanding of the operational context in which high-pressure washers are being employed. The sample comprised 9 male adult users who had considerable experience of using high-pressure washers ($M = 14.1$ yr; ranging from 6 to 20 yr). With each user, a contextual interview (Holtzblatt and Jones, 1993) of approximately 1 h was carried out. It included the observation of users while carrying out the task and covered a range of topics during the interview, such as mental model of the high-pressure washer and typical usage strategies. Mental models are conceptual representations of a technical system and its environment, which allow users to interpret system behaviour (Norman, 1988).

The analysis of the interview data revealed that only 2 out of 9 users had a good understanding of the complexity of the relationship between temperature and pressure levels and the kind of objects to be cleaned. Most users falsely believed that both temperature and pressure levels had a linear relationship with the effectiveness of the cleaning operation, independently of the object to be cleaned (i.e., the higher pressure and temperature levels, the better and quicker the object is being cleaned). However, the relationship described is often of a non-linear nature in that above certain settings (of which the precise level varies as a function of situational factors such as substance to be washed off) no additional effectiveness gains are made by further increases in parameter settings. Furthermore, users very frequently chose the maximum setting of water pressure whereas they typically opted for a medium setting for water temperature.

The results of the pilot study demonstrated that even among an experienced user population, there are misconceptions about the relationship between controls settings (temperature and pressure) and their effectiveness for certain tasks. This suggests that the conveyance of this critical information may be helpful to all users since even expert users appeared to have insufficient knowledge, let alone novice users.

3. Study I

3.1. Aim

The aim of study I was to determine the effectiveness of different designs of control labelling using a 2D paper prototype. Since the controls represent highly critical devices for influencing human–machine interaction, they may, if appropriately labelled, represent effective means of conveying information to users. Furthermore, the data obtained in this study with a paper prototype represented the first half for the subsequent comparison with the fully operational product employed in study II.

3.2. Method

3.2.1. Participants

Thirty male adult users took part in the study. They were selected on the basis of having some experience with the use of high-pressure washers (1–33 yr of experience). Most of them used high-pressure washers in a domestic context only (52.4%) while others (31.6%) employed them at work (e.g., caretaker, fire fighter) or in both contexts (16.0%). Participants did not receive any payment for their taking part in the study.

3.2.2. Design

A 3×7 within-subjects design was employed in the experiment, with the independent variables being control labelling and task scenario. Control labelling was varied at three levels (standard, substance-oriented labelling or

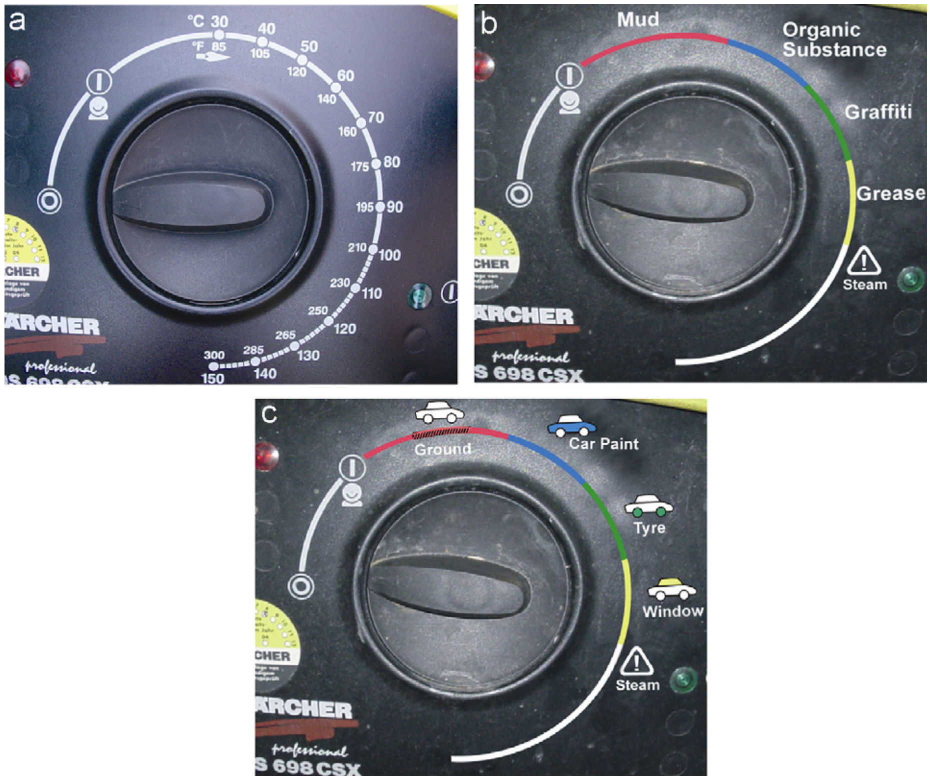

Fig. 1. Labelling of temperature control: (a) standard labelling, (b) substance-oriented labelling, (c) object-oriented labelling.

object-oriented labelling), with the labels being displayed in Fig. 1a–c. Substance-oriented labelling refers to an enhancement of the standard control label by providing the user with additional information about the best temperature setting for a given dirty substance (e.g., mud, paint). Object-oriented labelling provided information about the best temperature setting for a given object to be cleaned. Task scenario was manipulated at seven levels (front door, car paint, ground, motor, window, garden wall, car tyre), with some examples being given in Fig. 2. Type of labelling was only manipulated for the temperature control whereas the standard labelling for pressure display was used throughout (see Fig. 3). In the real system, the pressure display was not directly accessible due to a cover that had to be lifted by the user. It has to be noted that the recommendations given on the label are based on expert advice and have not undergone any empirical testing. The label was slightly modified for the purpose of the present experiment to avoid overlapping of settings recommended for different cleaning scenarios, which might have confused participants.

3.2.3. *Performance measures and user variables*

For the operation of a high-pressure washer, two of the most critical measures of user behaviour referred to the settings of *temperature* and *pressure*. Settings for both controls were chosen by the participants during the presentation of a series of task scenarios and recorded by the experimenter.

Since both controls are strongly associated with ecological performance of the high pressure washer, a German-language questionnaire of environmental concern was employed to measure different facets, such as water and energy consumption, recycling, and transport use (Schahn et al., 2000). The questionnaire has good psychometric properties (internal consistency: Cronbach's $\alpha = .93$; Schahn et al., 2000). For the present study, a shortened version was employed comprising 25 instead of 84 items.

In a post-experimental interview, an adapted version of the interview schedule from the contextual interview (see Section 2) was employed, with users being asked about their experience and habits of high-pressure washer usage. It also included verbal protocols of the users, explaining why control settings were chosen.

3.2.4. *Material and procedure*

Participants were presented with a series of seven cleaning scenarios in each of the three conditions. For each scenario they had to indicate the settings of temperature and pressure controls they would select to deal best with the object to be cleaned and the substance to be removed. Task scenarios were static and consisted of colour photographs of different cleaning tasks, of which some examples are shown in Fig. 2. The controls of the high-pressure washer were also presented to the participants by means of colour photograph. Participants chose the settings by marking a response sheet that showed a drawing of the controls. After having completed the ratings for the set of scenarios for one control label, they were

Fig. 2. Examples of tasks scenarios: (a) muddy car paint, (b) muddy car tyre, (c) mud on tiled ground.

Fig. 3. Pressure display.

given the same set of scenarios for the second control label and, finally, for the third. The order of presentation of the control labels was balanced out to control for carry-over effects. To control for order effects, an ANOVA was carried out, which showed that no such effects were observed ($F<1$). After having completed the ratings, participants were asked to fill in the environmental concern questionnaire. This was followed by a post-experimental interview, which aimed to gain an understanding of how users employed high-pressure washers in real life and, in particular, how they chose control settings.

3.3. Results

3.3.1. Temperature control setting

As the data in Table 2 show, users chose a wide range of temperature settings, reaching from 25.2 °C (scenario 'muddy car tyre' under substance-oriented labelling) to 106.8 °C (scenario 'front door graffiti' under standard labelling). It has to be noted that values higher than 100 °C could be chosen since the appliance was usable for steam cleaning. The analysis revealed that users selected lower settings when they were given guidance by substance- or object-oriented labelling. This difference was statistically significant between standard labelling and the two others ($F = 30.78$; df = 2,58; $p<.001$; LSD test: $p<.001$). The results also indicated that different settings were chosen as a function of the task scenarios presented ($F = 43.58$; df = 6,174; $p<.001$). The interaction between control labelling and task scenario was also significant ($F = 5.01$; df = 12,348; $p<.001$). This was due to differences in control settings across scenarios for the different labelling conditions, as the data in Table 2 demonstrate. Overall, the data show that users complied a little more strongly with the recommendations of substance-oriented labelling than of the object-oriented labelling.

3.3.2. Pressure control setting

In contrast to the temperature, there was no main effect of control labelling for the pressure parameter, as shown by the data in Table 2 ($F<1$). This is not surprising since the pressure label was not varied between experimental conditions. Interestingly, despite the absence of a label for the pressure control, users varied the setting as a function of cleaning scenarios ($F = 11.79$; df = 6,174; $p<.001$). There was no significant interaction ($F = 1.88$; df = 12,348; $p>.05$).

3.3.3. Pressure and temperature

A comparison of pressure and temperature controls management was carried out to test the assumption whether users treated both control parameters in a similar manner (i.e. a cleaning scenario that requires high temperature levels also requires high pressure levels), as the findings from the pilot study suggested. However, the correlation coefficient between pressure and temperature controls settings was not significant (for standard display: $r = .23$) and therefore failed to support the assumption.

3.3.4. Subjective ratings

For each display type separately, the correlation between environmental concern (water) and pressure as well as between environmental concern (energy) and temperature was examined. For the standard display, neither of them

Table 2
Setting of controls as a function of labelling and task scenario: recommended settings for substance-oriented labelling: mud (0–39 °C), organic substance (40–60 °C), graffiti (61–82 °C), grease (83–104 °C); recommended settings for object-oriented labelling: tiled ground (0–39 °C), car paint (40–60 °C), car engine (61–82 °C), car window 83–104 °C)

Scenarios	Standard labelling	Substance-oriented labelling	Object-oriented labelling	Overall
Temperature control setting (°C)	74.2	54.3	58.0	62.2
Front door graffiti	106.8[N/A]	79.7[N/A]	71.6[N/A]	86.0
Muddy car paint	51.7[N/A]	26.0*****	51.4****	43.1
Mud on tiled ground	46.8[N/A]	29.0*****	29.2****	35.0
Greasy car engine	102.8[N/A]	89.7**	73.7*****	88.8
Car window graffiti	98.3[N/A]	75.7****	87.1***	87.0
Moss-grown garden wall	68.6[N/A]	54.9[N/A]	59.0[N/A]	60.8
Muddy car tyre	44.3[N/A]	25.2*****	34.1[N/A]	34.5
Pressure control setting (bar)	94.2	93.4	93.3	93.6
Front door graffiti	108.9	103.1	102.7	104.9
Muddy car paint	70.3	73.8	71.2	71.8
Mud on tiled ground	97.3	100.6	104.1	100.7
Greasy car engine	77.2	85.4	86.3	83.0
Car window graffiti	86.1	87.3	86.6	86.6
Moss-grown garden wall	125.5	120.5	118.8	121.6
Muddy car tyre	94.3	83.3	83.1	86.9

Compliance rate (% of participants) with recommendation: *****90–100%, ****80–90%, ***70–80%, **60–70%, *50–60%; [N/A] no recommendation given on label.

had a significant influence on the setting of pressure or temperature, though the correlation was positive and slightly higher for pressure ($r_{\text{water}} = +.35$) than for temperature ($r_{\text{energy}} = -.06$). For the enhanced conditions (which was summarised across the two conditions), a similar picture emerged, with correlation coefficients pointing again into different directions ($r_{\text{water}} = +.44$; $p < .05$; $r_{\text{energy}} = -.21$; n.s.).

3.3.5. Post-experimental interview

The analysis of the post-experimental interviews revealed some difference between labelling conditions with regard to the extent to which the mental model guided their selection of the control setting. Under standard control labelling, users reported most frequently (80%) that they relied for at least one task scenario on their mental model for selecting control settings, compared to substance-oriented and object-oriented labelling (66.7% and 60.0%, respectively).

3.4. Discussion

The data analysis showed that the effect of control labelling on temperature settings was strong, demonstrating that enhanced labelling resulted in behavioural change. This indicated that the labels conveyed knowledge to the user effectively, that is, all four phases of Wogalter's (1999) information processing model (i.e. behavioural advice needs to be *noticed* and *understood* by users, it needs to *convince* users so that, finally, they will *implement* it) were successfully passed through. The effectiveness of the control labelling may be increased in a paper prototype since it may be less likely that communication breaks down in any of these phases, compared to a real appliance. Particularly the phases *taking notice* and *implementation* are less critical in a paper prototype than a fully operational product since there are fewer information sources competing for the user's attention (i.e. taking notice is facilitated) and less effort is required to act on the basis of the advice given (i.e. implementation is facilitated).

The analysis revealed that substance-oriented labelling had a similar compliance rate like object-oriented labelling. This suggests that both labels are similarly effective in modifying user behaviour. However, which of the two is chosen is not without consequence since each type of labelling refers to distinct functions. Substance-oriented labelling is primarily concerned with the cleansing function, that is, controls settings need to be chosen such that the undesired substance is being washed off (i.e. usually a minimum required setting, though there may be exceptions). Object-oriented labelling refers to the protective function, that is, the controls have to be set such that no damage is made to the object to be cleaned (maximum allowable setting). There are circumstances in which a difficult trade-off may be required (e.g., if the settings required to wash off a substance may do damage to the cleaned object). Clearly, more work needs to be done to create sensible categories that subsume all possible cleaning scenarios of a high-pressure washer, taking into account the sometimes conflicting requirements of the cleansing and protective function. This category system is likely to entail a combination of substance- and object-oriented scenarios, perhaps analogous to the cleaning scenarios of a washing machine.

If no labelling is available as in the standard condition, the selection of the most appropriate control settings has to be guided by the user's mental model alone. Generally, the user's mental model was less accurate than the information conveyed by the enhanced control labelling, as indicated by the more optimal temperature settings in the enhanced control conditions. However, the results showed that users adjusted their temperature and pressure levels according to the requirements in each cleaning scenario. From this strong variation in temperature and pressure settings, it may be concluded that users considered both parameters to be important for carrying out an effective cleaning job. The active management of power controls as seen in the present study is not a commonly observed phenomenon in domestic appliances, as other research has demonstrated (Sauer et al., 2004). In that study, users operating a fully operational prototype of a vacuum cleaner made little use of suction control to respond to different cleaning tasks. The difference in the frequency of control actions may be due to appliance- and task-specific effects (i.e. the power controls of some appliances are more important than others due to the nature of the task to be completed) or may be moderated by prototype fidelity (i.e. paper prototypes facilitate control actions). Which of the two alternatives explanation represents the stronger one may be answered in study II, wherein a fully operational appliance was being used.

In contrast to the findings of the pilot study, there was little evidence in the data that users treated temperature and pressure as similar parameters. This suggests that the assumption derived from the pilot study (which proposed "the higher the setting, the more effective the cleaning process") cannot be maintained in its general form. While in the pilot study users made a general statement, in study I different cleaning scenarios were presented, requiring a more situation-specific judgement involving the mechanical and thermal robustness of the object to be cleaned. The objects in scenarios "greasy car engine" and "car window graffiti" were two examples that were considered to be rather heat resistant but mechanically more delicate.

4. Study II

4.1. Aim

The aim of study II was to examine the effectiveness of two different control labels using a fully operational high-pressure washer. This study provides the second half of the data, allowing for a comparison of user behaviour with paper prototypes and fully operational products, which will be presented in Section 5.1.

4.2. Method

4.2.1. Participants

Thirty-four participants (23.5% female) took part in the study. Their ages ranged from 16 to 66 years ($M = 29.6$ yr). Only participants who had some experience of high-pressure washer usage were selected for the study (2–24 yr of experience). They were paid €15 for their participation.

4.2.2. Design

A 2×4 mixed design was employed in the experiment, with the between-participants variable *labelling of control panel* and the within-participants variable *task scenario. Labelling of control panel* was varied at two levels (standard, enhanced). Both labels were modelled on the ones used in study I, with the standard label being presented in Fig. 1a and the object-oriented label (see Fig. 1c) being selected as an enhanced control label. The object-oriented label was chosen rather than the substance-oriented label for two reasons. First, users found the former easier to use and, second, the protective function (displayed by object-oriented label) was more critical than the cleansing function (displayed by substance-oriented label). The task scenarios employed were also modelled on the pictures presented in study I. The variable *task scenario* was manipulated at 4 levels (tiled ground, car paint, car window, car tyres).

4.2.3. Performance measures

Six measures of performance were taken during the experimental trials: settings of temperature (°C) settings of pressure controls (bar), energy consumption (kWh), water consumption (L), task completion time (s), and achieved cleanliness. The participants' setting of the temperature and pressure controls was measured through observation by one of the experimenters. Energy consumption, water consumption and task completion time were measured by an electricity metre, a water metre and a stopwatch, respectively. Achieved cleanliness was assessed by independent ratings of the two experimenters, using a three-point scale describing standards of cleanliness (entirely satisfactory, partly satisfactory, and unsatisfactory). The assessment was based on a scoring sheet that provided qualitative descriptions of the different standards of cleanliness that may be observed (e.g., for partly satisfactory standard: isolated patches of white paint are still left on the ground).

4.2.4. User variables

In addition to the behaviour observed in the experiment, we sought to complement the database by collecting several user variables. To measure *environmental concern*, the same questionnaire as in the study I was employed (for details, see above). Also, the post-experimental interview schedule from study I was used again to measure user's usual behaviour and to identify the reasons why certain control actions were taken.

4.2.5. Material and procedure

A second-hand car (Mitsubishi Colt) was purchased for the experiment. Before being used for the experimental work, its exterior was very thoroughly cleaned. Prior to each experimental trial, two substances (mud and paint) were applied to the car and its surroundings following a strict experimental schedule. A fixed amount of mud (2.5 L) and paint (0.2 L) was applied with paintbrushes onto the car in a standardised form, as shown by four highresolution photographs that had been taken to facilitate the standardisation of the experimental conditions.

The participants were asked to clean a car and its surroundings with a high-pressure washer (model Kärcher HDS 698CSX). They were told that they should carry out the tasks as if they were to clean their own car. This was followed by a brief introduction to the technical features and operation of the high-pressure washer.

The cleaning task consisted of the following four parts (which represented the different levels of the independent variable "cleaning scenario"): *tiled ground* (covered with mud), *car paint* (covered with mud), *car tyre* (covered with mud), and *car window* (graffiti). These scenarios were modelled on the static scenarios employed in study I (see Fig. 2). The order of presentation was fixed to minimise the impact of preceding tasks on subsequent tasks (e.g., removing graffiti from the car windows would have also washed off some of the mud on the ground). After having completed the cleaning tasks, participants were interviewed as in the previous study to collect data about their experience of high-pressure washer usage and environmental concern.

4.3. Results

4.3.1. Temperature control setting

The means displayed in Table 3 showed that there was no main effect of control labelling ($F<1$). However, the data indicated that different settings were chosen as a function of the task scenarios presented ($F=23.8$; df $=3{,}90$; $p<.001$). Furthermore, the interaction between control labelling and task scenario was significant ($F=21.5$; df $=3{,}90$; $p<.001$). This was because under standard labelling, there was little variation in settings chosen (range: 3.3) whereas variation was drastically increased under enhanced labelling (range: 51.3).

4.3.2. Pressure control setting

As observed for temperature control, there was no main effect of control labelling for the pressure parameter, shown by the data in Table 3 ($F<1$). Interestingly, despite the absence of a label for pressure control, users varied the setting as a function of cleaning scenarios ($F=3.2$; df $=3{,}96$; $p<.05$). No interaction was observed ($F<1$).

4.3.3. Energy consumption

This measure refers to the electricity consumption (kWh) of the motor. This measure was not taken at the end of each scenario. Therefore, only a one-way analysis was carried out on the overall trial consumption score. The data are presented in Table 3, showing no difference as a function of control labelling ($F<1$).

4.3.4. Water consumption

This concerns the amount of water (L) that was consumed during the trial. As above, this measure was not taken separately for each scenario but for the entire trial. As the data in Table 3 showed, there was little difference between labelling conditions ($F<1$).

Table 3
Performance measures as a function of controls labelling and task scenario; recommended settings for object-oriented labelling: Mud on tiled ground (0–39 °C), muddy car paint (40–60 °C), muddy car tyre (61–82 °C), car window graffiti (83–104 °C)

	Standard labelling	Enhanced labelling	Overall
Temperature control setting (°C)	55.7	55.8	55.9
Mud on tiled ground	55.5[N/A]	30.0*****	44.3
Muddy car paint	56.0[N/A]	51.1**	53.7
Muddy car tyre	53.9[N/A]	60.6**	57.1
Car window graffiti	57.2[N/A]	81.3**	68.5
Pressure control setting (bar)	62.9	66.9	64.8
Mud on tiled ground	60.3	59.4	59.9
Muddy car paint	62.8	70.9	66.6
Muddy car tyre	62.2	65.9	64.0
Car window graffiti	66.1	71.3	68.5
Energy consumption (kWh)	.239	.233	.236
Water consumption (L)	36.96	37.15	37.06
Duration (s)	512.8	488.0	500.4
Achieved cleanliness (1–3)	2.40	2.34	2.37
Mud on tiled ground	2.17	2.19	2.18
Muddy car paint	2.47	2.44	2.45
Muddy car tyre	2.39	2.25	2.32
Car window graffiti	2.56	2.50	2.53

Compliance rate (% of participants) with recommendation: *****90–100%, ****80–90%, ***70–80%, **60–70%, *50–60%; [N/A] no recommendation given on label.

4.3.5. Task completion time

This measure refers to time (s) needed to complete the total task scenario. Again, no data were collected as a function of task scenario. As the data in Table 3 showed, users needed 8–9 min for all four task scenarios together. Again, there was little difference between labelling conditions ($F<1$).

4.3.6. Achieved cleanliness

This measure provided a cleanliness rating given by the experimenters on a three-point scale (ranging from 1 = *unsatisfactory* to 3 = *entirely satisfactory*) are presented in Table 3. It shows that overall cleaning performance was considered to be of high quality. No significant difference was found between labelling conditions ($F<1$). There was an effect of scenario type ($F=6.4$; $\mathrm{df}=1,3$; $p<.001$), with objects like 'car paint' and 'car window' being cleaned more thoroughly than the two others. No interaction was observed ($F<1$).

4.3.7. Environmental concern and user behaviour

The pattern observed here was very similar to the one found in study I. The correlation between temperature setting and environmental concern (energy) was $r=-.38$ for the enhanced labelling and $r=-.14$ for the standard labelling. For the pressure setting and environmental concern (water), it was $r=+.29$ (enhanced) and $r=+.04$ (standard), respectively. While none of these correlation coefficients obtained significance (owed to the small sample size in comparison to a typical correlation study), it is interesting to note that the pattern of study I (i.e. positive correlation for water and negative correlation for energy) could be replicated.

4.4. Discussion

The results did not show a significant main effect of control labelling. This was due to the recommendations given on the label, which increased mean temperature settings for some scenarios and decreased it for others, compared to the standard condition. Overall, a majority of users followed the recommendations for each condition. However, the data suggested that for scenarios requiring a high temperature setting (e.g., car window), users showed a tendency towards selecting a medium setting rather than strictly following these recommendations.

The most interesting finding was that under enhanced control labelling users made bigger changes to the temperature settings than under standard control when faced with a new scenario. This suggests that the enhanced control label induced changes in that it conveyed information to the user about recommended settings, which was not available to the user in the standard condition. It may be speculated that under the standard labelling condition, users may have been unsure about what setting to choose in that scenario so that they opted for a medium control setting to see whether it would be effective in cleaning the object. When operating the real appliance, users always observed (even if the setting was far from being optimal) some positive effect in the experimental scenarios in that the substance was (at least slowly) being washed off. Therefore, they were likely to have persevered with that setting because they received no feedback about it being sub-optimal. This would correspond to the satisficing heuristic, which states that humans would accept an option that met their minimum requirements and no better option was available at that time (Simon, 1978). Only if there had been no effect at all on the substance to be cleaned, users would have been likely to take some action by changing the setting. Clearly, these explanations are somewhat speculative, as they cannot be substantiated by the data of the present experiment.

The examination of the pattern for the real appliance suggests a distinctly different use of temperature controls in the enhanced labelling condition compared to the standard condition. This appears very different from the pattern observed in the paper prototype condition (where no feedback of the effectiveness of a chosen setting was available) and represents perhaps the most important finding of this study. This is formally tested in a comparative analysis in the subsequent section and will hence also be discussed there.

There was evidence that users changed the settings of pressure controls as a function of the cleaning scenario although no label guided them. This is even more remarkable if one considers that pressure control was not directly accessible because of a cover so that gaining access incurred costs in terms of physical effort and additional time requirements. This emphasises the importance of pressure control for task completion since even under difficult circumstances (i.e. system design that required increased effort expenditure), users made regular adjustments of pressure levels.

With regard to the association of different facets of environmental concern and settings of control, the pattern found in study I could be replicated in study II. This supports the assumption that users scoring high on environmental concern for energy tended to choose lower temperature settings (correlation coefficients in both studies between $-.38$ and $-.14$) whereas higher pressure levels were chosen for users scoring high on environmental concern for water (correlation coefficients in both studies between $+.44$ and $+.04$). The strength of the association typically found in studies correlating environmental concern with behaviour is around $r=.35$ (Hines et al., 1986). Of particular interest was the unexpected positive correlation observed for water. This may have been different because in Northern and central European countries energy represents a more precious resource than water. A different pattern may have been found in an Arabic country with an abundance of energy but a shortage of water. An alternative explanation may be that water pressure provides a more visible feedback to the user whether an effective setting has been chosen than temperature in term of washing off undesired substances from objects.

5. Comparative analysis of paper prototype and fully operational appliance

5.1. Aim

This section aims to examine the suitability of a 2D paper prototype for modelling 3D products. This refers to the first goal of this article addressing an important methodological question of the utility of prototypes of lower fidelity.

5.2. Method and analysis

In order to address that question, the data sets of the two studies have been merged. Since the data sets differed in size (i.e. number of participants) and structure (i.e. control labelling was used as a within-subjects variable in Study I but as a between-subjects variable in Study II), some modifications of the data sets were required before they could be merged. The first step was to exclude the data points that were taken under the condition ''substance-oriented labelling'' because this condition was not replicated in study II (see Section 4.2.2 for an explanation). The second step was to exclude the data for the task scenario conditions 'moss-grown garden wall', 'front door graffiti' and 'greasy car engine' because there were no equivalent conditions in Study II (these could not be integrated into the cleaning scenario of Study II).

These two steps matched the experimental conditions of study I to those of study II. The third step was to select at random 15 participants from study I of whom the data points from the ''standard control'' condition were taken and, from the remaining 15 participants from Study I, the data points from the ''object-oriented labelling'' condition were used. This was to transform the within-subjects design into a between-subjects design and to ensure that the means of the two conditions were uncorrelated. *t*-Tests were carried out to ensure that the means of the newly created subsets of the sample were not statistically different from the original sample ($p>.05$ for all experimental conditions). The conversion of the within-subjects design into a between-subjects design was considered to be a conservative procedure since it reduced the statistical power of the test.

A $2\times2\times4$ mixed design has resulted from this procedure, with a total sample of $N=64$. Contextual fidelity was a between-subjects variable and was varied at two levels: low ($N=30$) vs. high ($N=34$). At the low-fidelity level a paper prototype was used together with photos of the soiled object to be cleaned whereas at the high-fidelity level a real high-pressure washer was used on an artificially soiled car. Type of control labelling (standard vs. enhanced) was also a between-subjects variable while task scenario was a within-subjects variable being varied at 4 levels (tiled ground, car paint, car window, car tyre).

5.3. Results

The analysis focused on the parameter ''temperature control setting'' since this efficiency measure was collected in both studies because temperature control labelling was manipulated as an independent variable. A three-factorial analysis of variance was carried out on the data set. The results are presented in Fig. 4. It shows that there were no main effects of fidelity ($F<1$) and labelling ($F<1$). However, there were considerable variations in temperature settings as a function of cleaning scenario, confirmed by a highly significant main effect ($F=52.5$; df $=3{,}180$; $p<.001$). Most interesting is clearly the 3-way interaction between all independent variables ($F=2.7$; df $=3{,}180$; $p<.05$). As the data in Fig. 4 show, there is a stronger variation in settings across the four different cleaning scenarios for paper prototypes (i.e. for standard and enhanced labelling together; $SD_{low}=23.8$) than for the real appliance (i.e. for standard and enhanced labelling together; $SD_{high}=11.8$). As can also be seen, this difference in variation is larger under standard display ($SD_{low}=25.9$; $SD_{high}=1.9$) than under enhanced display ($SD_{low}=25.0$; $SD_{high}=17.8$), resulting in the significant interaction described. This indicates that the moderating effect of fidelity decreases with increasing information richness of the object under investigation (e.g., in the form of enhanced labelling).

In addition to this 3-way interaction, two significant 2-way interactions were observed. First, the interaction fidelity x scenario confirms that the paper prototype scenarios show higher variations as a function of task scenarios than real usage scenarios ($F=15.1$; df $=3{,}180$; $p<.001$). Second, the interaction display x scenario shows that there is a stronger variation in setting selection for enhanced displays than for standard ones ($F=3.8$; df $=3{,}180$; $p<.01$). No other significant interaction was observed.

An analysis was carried out to determine the frequency with which recommendations on the control labels were complied with. The results showed a higher compliance rate for the paper prototype than for the real high-pressure washer (see Fig. 4). Finally, the analysis revealed that within each condition, the variance of settings chosen by users was larger for standard labelling than for enhanced labelling.

5.4. Discussion

There are two important points that emerged from this comparative analysis. First, enhanced control labelling had a strong influence on user behaviour for both paper prototypes and the real system because users changed the temperature setting after each scenario, largely following recommendations of the label. Second, under standard

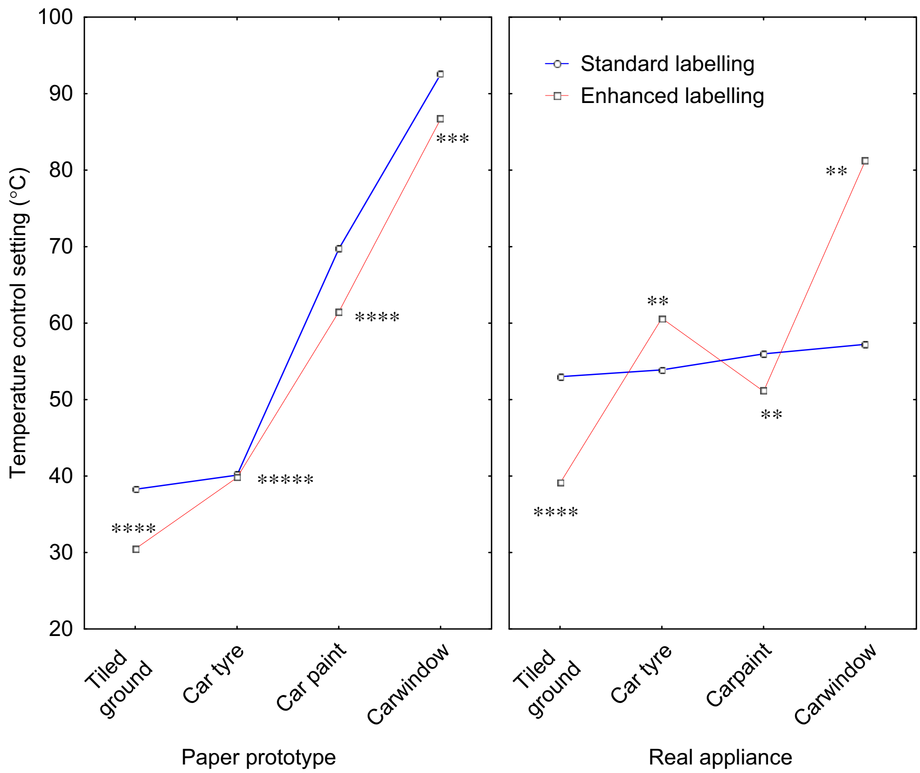

Fig. 4. Temperature setting as a function of fidelity level, type of labelling and scenario type; mud on tiled ground (0–39 °C), muddy car paint (40–60 °C), muddy car tyre (61–82 °C), car window graffiti (83–104 °C); Compliance rate (% of participants) with recommendation: *****90–100%, ****80–90%, ***70–80%, **60–70%, *50–60%.

control labelling users showed considerable variation in temperature settings across task scenarios in the paper prototype simulation only, whereas in the real usage scenarios few changes to the settings were observed. Three hypotheses are put forward to explain this interaction: *Feedback hypothesis*, *habit hypothesis*, and *effort hypothesis*. These are now considered in turn.

5.4.1. Feedback hypothesis

There are two types of feedback available in the different experimental conditions: *labelling-based* and *impact-related*. While the former is linked to the information provided by enhanced labelling, impact-related feedback is associated with the real appliance by giving the user an indication of the degree to which the chosen control setting was operationally effective (i.e. dirt is being washed off). Therefore, there were considerable differences between the four experimental groups with regard to the kind of feedback provided: *real appliance with enhanced labelling* (both types of feedback), *paper prototype with standard labelling* (no feedback), *real appliance with standard labelling* (impact-related feedback only), and *paper prototype with enhanced labelling* (labelling-based feedback only). When labelling-based feedback was available, the user followed this recommendation, as indicated by remarkably high compliance rates. When there was no feedback available (i.e. paper prototype with standard labelling), the users needed to make use of their knowledge of the appliance to select appropriate temperature settings in response to the considerable variability in operational requirements across task scenarios. When only impact-related feedback was available (i.e. real appliance with standard labelling), many users maintained the setting first chosen throughout the trial rather than trying to optimise their chosen setting by means of iterative improvements. Maintaining the first setting may be indicative of an approach that resembles a satisficing strategy (Simon, 1978), which suggests that users choose and maintain the first available option that meets their minimum requirements. Generally, impact-related feedback is more ambiguous than labelling-based feedback. Although impact-related feedback provides an indication whether the chosen setting works (i.e. dirt is being washed off), it is difficult for users to judge whether a different setting would be even more effective (i.e. dirt is being washed off more quickly).

5.4.2. Habit hypothesis

Labelling-based information provides the possibility to choose a more optimal setting, of which users in both conditions took advantage. In the absence of guidance through labelling-based information, users in highly familiar settings are less likely to break with habits and routines that normally guide their task management process and which have been established over a longer period of time. There is evidence from analogous technical system such as the vacuum cleaner that users generally make few changes to power control settings (Sauer et al., 2002), which is a pattern that has also been observed in the present study. In contrast to the real appliance, breaking the habit is facilitated in the more unfamiliar context of "operating" a paper prototype. Similarly, research in laboratory-based decision-making found that increased unfamiliarity of the context is more likely to make humans change their routines and habits (Betsch et al., 2001). Since it is not possible to recur to habitual behaviour, users are obliged to make knowledge-based judgements in the task scenarios, that is, they have to rely on their understanding of the system and the task to select appropriate temperature settings.

5.4.3. *Effort hypothesis*

There is generally less physical effort required to change settings of paper prototypes than of real systems. In the case of a high-pressure washer, the user is required to make bodily movements beyond the manipulation of controls (e.g. walking a few metres to reach distant controls on main unit of high-pressure washer, bending forward and opening a cover to gain access to pressure control) since some controls were in a distal rather than proximal position in relation to the user. Previous research has demonstrated that increasing the distance between user and controls leads to reduced frequency of user intervention (Sauer et al., 2002), as larger distances require more physical effort expenditure. Under enhanced control labelling, the provision of action-related information may have increased user motivation, which resulted even with the real high-pressure washer in more frequent manipulations of the control settings being carried out.

On the basis of the available data, it is not possible to assess the relative influence of each of the three factors (termed hypotheses here since they represent assumptions that can be empirically tested in future studies). Future research may be able to determine the respective impact of these factors, which would provide a better understanding of the effects of conducting usability tests at different fidelity levels.

6. Main discussion

The present work provided some evidence for the effectiveness of enhanced control labelling to improve the efficiency and environmental friendliness of product usage. Clearly, there is more work needed to create suitable categories of cleaning scenarios, which users can choose from. While these categories may represent a combination of substance- and object-related categories, object-related categories are likely to dominate since they represent the more critical factor due to their protective function.

Concerning the influence of prototype fidelity in consumer product design, the results showed that the real appliance provided a somewhat different picture of user behaviour than the paper prototype, suggesting the need for caution in interpreting behavioural data obtained with a paper prototype.

The finding that there is an influence of fidelity appears to be, at first, in contrast to most studies from the research literature. However, a closer look at the studies reveals several differences compared to the current work (see also Table 1). First, the present work examined efficiency measures rather than merely looking at usability errors. The relationship between usability errors and efficiency measures is complex. Efficiency measures represent global indices and are influenced by a number of factors, with the usability error representing merely one of these. There are factors beyond usability errors that influence efficiency measures (e.g., setting goals for cleaning standards), which may lead to changes in system operation that result in increased or decreased efficiency without any usability error occurring. Because of the many factors influencing the efficiency measure, it may not be surprising that different fidelity levels have a different impact on usability errors than on efficiency measures. Second, our work modelled aspects of a 3D product by means of a 2D prototype. This is clearly a wider gap to bridge than when the prototype has the same number of dimensions as the target product. The explanations that have been given above may also provide indications of why 2D and 3D prototypes provided somewhat different results. Despite the differences found in user behaviour as a function of prototype fidelity, the results do not imply that 2D prototypes are not suitable. The important point is that the designer needs to be aware that paper prototypes may generally produce stronger effects when comparing different design options than fully operational prototypes.

In addition, the concept of prototype fidelity should not be discussed in isolation but needs to be considered in relation to other aspects of fidelity that refer to a wider usage context. This includes the fidelity of *user characteristics*, *task scenario*, and *physical and social environment*. Clearly, the four dimensions of fidelity are not independent of each other. Choosing a prototype of a certain fidelity level places constraints on the selection of task scenarios, the social and physical environment, and, indeed, the user. For example, the actual complexity of real cleaning scenarios can only be modelled when a fully operational prototype is employed to clean a real object while, in contrast, a 2D prototype would only allow the modelling of a pure decision-making task (e.g., user would inform experimenter about what power controls setting is to be chosen), similar to the information display board paradigm used by Verplanken et al. (1997). This implies that in the studies reviewed, it was not only the fidelity of the prototype that differed but also the levels of other fidelity dimensions. Therefore, there may be a risk of the effects of prototype fidelity being to some extent confounded with the effects of other fidelity dimensions. This issue needs to be addressed by future research. Overall, all these deliberations suggest a need for a more all-encompassing view on fidelity that goes beyond the issue of prototype fidelity.

Acknowledgements

We gratefully acknowledge the financial support of the German Research Foundation (DFG) for carrying out this work (Research Grant: SFB392/TFB55). We would also like to thank Dr H Gehringer and Mr A Leschtar from Alfred Kärcher GmbH & Co. KG for providing us with advice and technical support.

References

Betsch, T., Haberstroh, S., Glöckner, A., Haar, T., Fiedler, K., 2001. The effects of routine strength on adaptation and information search in recurrent decision making. Org. Behav. Hum. Dec. Process. 84, 23–53.

Catani, M.B., Biers, D.W., 1998. Usability evaluation and prototype fidelity: users and usability professionals. Proceedings of the Human Factors Society 42nd Annual Meeting, pp. 1331–1335.

Chapanis, A., 1996. Human Factors in Systems Engineering. Wiley, New York.

Cox, E.P., Wogalter, M.S., Stokes, S.L., Tipton Murff, E.J., 1997. Do product warnings increase safe behavior? A meta-analysis. J. Pub. Policy Marketing 16, 195–204.

Elliot, L.R., Dalrymple, M.A., Schifflett, S.G., Miller, J.C., 2004. Scaling scenarios: development and application of C4ISR sustained operations research. In: Schiflett, S.G., Elliot, L.R., Salas, E., Coovert, M.D. (Eds.), Scaled Worlds: Development, Validation, and Applications. Ashgate Publishing Limited, Aldershot, pp. 119–133.

Hall, R.R., 1999. Usability and product design: a case study. In: Jordan, P., Green, W.S. (Eds.), Human Factors in Product Design. Taylor & Francis, London, pp. 85–91.

Hines, J.M., Hungerford, H.R., Tomera, A.N., 1986. Analysis and synthesis of research on responsible environmental behavior: A meta-analysis. J. Environ. Educ. 18 (2), 1–8.

Holtzblatt, K., Jones, S., 1993. Contextual inquiry: A participatory technique for system design. In: Schuler, S., Namioka, A. (Eds.), Participatory Design: Principles and Practices. Erlbaum, Hillsdale, pp. 177–210.

Hsu, S.H., Peng, Y., 1993. Control/display relationship of the four-burner stove: a re-examination. Hum. Factors 35 (4), 745–749.

Jordan, P.W., 1998. An Introduction to Usability. Taylor & Francis, London.

McCarthy, R.L., Ayres, T.J., Wood, C.T., Robinson, J.N., 1995. Risk and effectiveness criteria for using on-product warnings. Ergonomics 38, 2164–2175.

Nielsen, J., 1990. Paper versus computer implementations as mock-up scenarios for heuristic evaluation. Proceedings of the IFIP INTERACT90 Third International Conference Human–Computer Interaction (Cambridge, UK, August 27–31), pp. 315–320.

Norman, D.A., 1988. The Psychology of Everyday Things. Basic Books, New York.

Rooden, M.J., 1999. Prototypes on trial. In: Jordan, P., Green, W.S. (Eds.), Human Factors in Product Design. Taylor & Francis, London, pp. 85–91.

Rudd, J., Stern, K., Isensee, S., 1996. Low vs. high-fidelity prototyping debate. Interactions 3 (1), 76–85.

Säde, S., Niemenen, M., Riihiaho, S., 1998. Testing usability with 3D paper prototypes—Case Halton system. Appl. Ergon. 29, 67–73.

Sanders, M.S., McCormick, 1993. Human Factors in Engineering and Design. McGraw-Hill, New York.

Sauer, J., Wiese, B.S., Rüttinger, B., 2002. Improving ecological performance of electrical consumer products: the role of design-based measures and user variables. Appl. Ergon. 33 (4), 297–307.

Sauer, J., Wiese, B.S., Rüttinger, B., 2003. Designing low-complexity electrical consumer products for ecological use. Appl. Ergon. 34 (6), 521–531.

Sauer, J., Wiese, B.S., Rüttinger, B., 2004. Ecological performance of electrical consumer products: the influence of automation and information-based measures. Appl. Ergon. 35, 37–47.

Schahn, J., Damian, M., Schurig, U., Füchsle, C., 2000. Konstruktion und evaluation der dritten version des skalensystems zur erfassung des umweltbewusstseins (SEU-3). Diagnostica 46 (2), 84–92.

Sefelin, R., Tscheligi, M., Giller, V., 2003. Paper prototyping—what is it good for? A comparison of paper- and computer-based low-fidelity prototyping. CHI '03 Extended Abstracts on Human factors in Computer Systems, April 2003.

Simon, H.A., 1978. Rationality as process and as product of thought. Am. Econ. Rev. 68, 1–16.

Verplanken, B., Aarts, H., van Knippenberg, A., 1997. Habit, information acquisition, and the process of making travel mode choices. Eur. J. Soc. Psychol. 27 (5), 539–560.

Virzi, R.A., Sokolov, J.L., Karis, D., 1996. Usability problem identification using both low- and high-fidelity prototypes. Conf. Proc. Hum. Factors Comput. Syst.: CHI 96, 236–243.

Walker, M., Takayama, L., Landay, J., 2002. High-fidelity or low-fidelity, paper or computer medium? Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting: HFES2002, pp. 661–665.

Wenzel, H., Hausschild, M., Alting, L., 1997. Environmental Assessment of Products, vol. 1. Chapman & Hall, London.

Wiese, B.S., Sauer, J., Rüttinger, B., 2004. Consumers' use of written product information. Ergonomics 47 (11), 1180–1194.

Wiklund, M., Thurrot, C., Dumas, J., 1992. Does the fidelity of software prototypes affect the perception of usability? Proceedings of the Human Factors and Ergonomics Society 36th Annual Meeting: HFES1992, pp. 399–403.

Wogalter, M.S., 1999. Factors influencing the effectiveness of warnings. In: Zwaga, H.J.G., Boerseman, T., Hoonhout, H.C.M. (Eds.), Visual Information for Everyday Use: Design and Research Perspectives. Taylor & Francis, London, pp. 93–110.