1 The functioning of products when in usage, a graphical representation

An important component in the IO curriculum is the design of serviceable products. Such design would not be possible without some idea, however (partially) implicit it may be, of the functioning of this future product, the conditions, the user, the manner and the intended result.
In many cases the use is the culmination of a product design.
To get a clear picture of the role of the intended use/user, this chapter provides a graphical depiction of the functioning of a product when in use. To an important extent this depiction, or representation, is inspired by findings in empirical research (see Kanis, 1998).
In this first chapter we start the description of product use with, on the one hand, a technical artefact – in other words a product – and, on the other hand, a user who is using the product in question in a certain environment. In this context we discuss the concept of 'interaction' and the influence of more general human characteristics in this interaction. We will conclude this chapter with a description of the concept of 'usability', followed by a number of the theoretical considerations regarding the proposed depiction of the user-product interaction.

1.1 The functioning of a consumer product: technical-physical aspects

Figure 1.1 reflects the functioning of a product as a concomitance of the following elements:

- The product in question, as determined by the product characteristics and the resulting functionalities. A functionality is the product's ability to function in the protection, support and replacement of human activities, and/or in the expansion of human capabilities.
- Use actions, being any physical action that activates or utilises the functionality of a product: actions including posture, foot movements, speech, sometimes even eye movements. Often a use action will result in physical contact with a product (component), but this is not necessarily always the case. Using a hand to shield the eyes from the sun when trying to see the screen of an ATM machine is part of a use action. Likewise, head movements for the purpose of reading a road sign, for instance, can be considered use actions.
- Environment-related factors insofar as they directly affect or interact with a product; for instance temperature, (uneven) surfaces etc.

The concomitance of these three factors determines the functioning of a product with respect to:

- product performance, e.g. the resulting functionality of a product, and
- side effects during the functioning (e.g. vibration, noise, (the risk of) accidents etc.; in other words, effects that are undesirable in most cases but not always (sometimes these can serve as feedback).

Figure 1.1 illustrates a purely technical-physical phenomenon, also with respect to use actions – consider, for instance, actions (re)produced with the use of a robot. Thus the role of use actions

in figure 1.1 remains limited to the impact they have on the functioning of the product. As soon as the *origin* of use actions comes into play, this definitely involves human behaviour, see next paragraph.

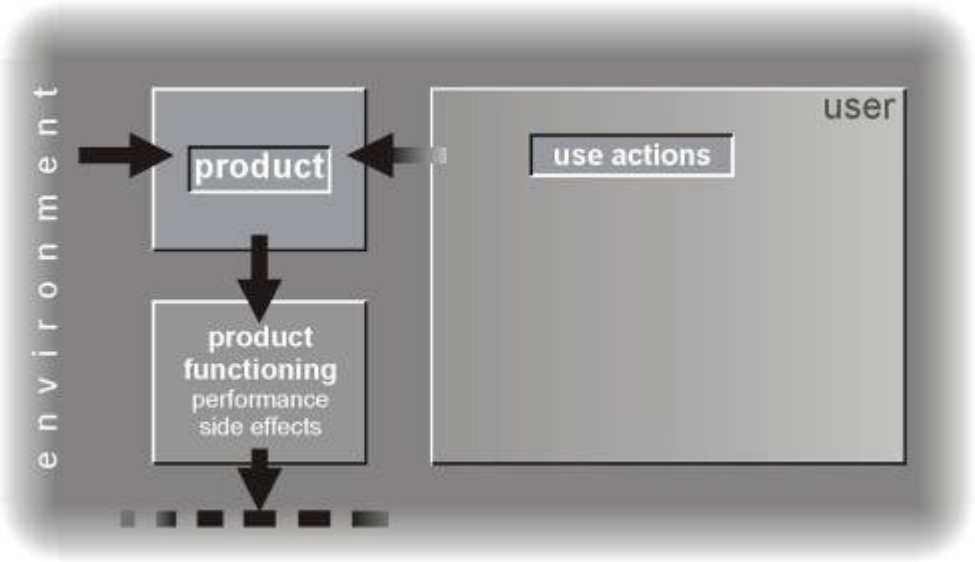


Figure 1.1 *The functioning of a product as a result of a technical-physical process, namely the concomitance of a product and use actions in an environment*

The product in figure 1.1 looks rather diagrammatical. For instance, why is the side where the use action arrow comes in not specified as ‘controls’? You could do this, except you would have to interpret ‘control’ so broadly that, for instance, an entire component of a product is a ‘control’, such as a handle, or virtually an entire product – consider a glove or a pen. Rather contrived, therefore.

In a similar manner, on the side of the outgoing function arrow the product could have been specified as ‘display’, but the same objections would arise; consider, for instance, the position in space of a product that is being held, such as a knife aimed at a body part. Furthermore, the information being sought about the functioning need not be associated with a product at all, or only partially. Consider all those products that work on something in the environment (corkscrews, screwdrivers, vacuum cleaners, irons, remote controls and so on).

1.2 User involvement: use activities

☞ Figure 1.2 is an expansion of figure 1.1 with regard to use activities: in addition to use actions these are the preceding observation (perception) and any reasoning and interpretations (cognition), including the effort (mental, physical) associated with any use activity as experienced by the user.

With regard to perceptions (seeing, hearing, feeling, smelling, ...) figure 1.2 may relate to:

- external product characteristics that can be ‘read’, such as shape, dimensions, weight, colour, design, texture, transparency, and structure/grouping of parts,
- observable functioning: product performance, side effects, and
- any relevant environment characteristics.

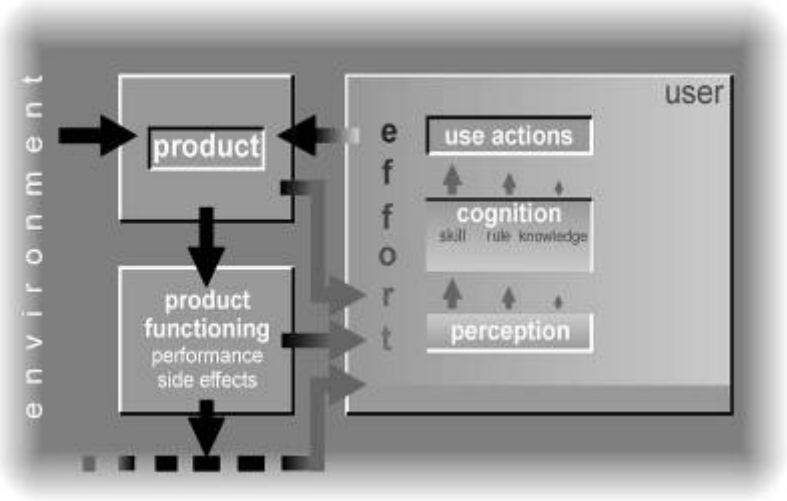


Figure 1.2 The functioning of a product under the influence of use activities

From a theoretic viewpoint, it is difficult to distinguish between perceptions and cognitions (Neisser, 1976). But in practice this need not be so difficult. People may not have a problem distinguishing between what they have observed and whether they have drawn any conclusions from this or looked for some meaning behind it. Compare the example in figure 1.3.
As discussed more extensively later in this document, it makes quite a difference to a designer who wants to convert use observations into product improvements whether a certain cue was not noticed or not understood, or interpreted differently from the way it was intended.

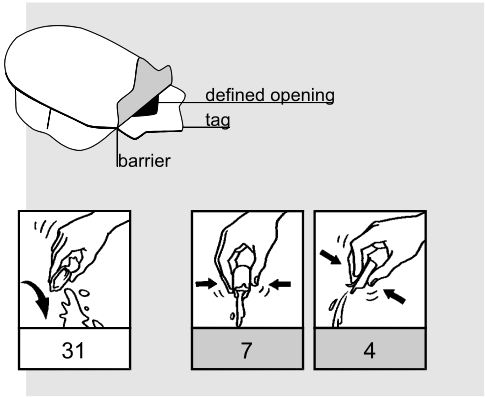


Figure 1.3 The use of a newly designed coffee creamer tub. On the right the tub has not been opened any further than intended in the design, namely to the extra strong glue line between the tub and the foil, followed by a dosing as intended by the designers, in other words squeezing between the two sides or between the top and bottom (in only approx. 15% of cases).
On the left by far the most frequently occurring use in practice: the foil is pulled off completely, see the following excerpt from an interview with a test subject.

[question] "Did you notice anything unusual when opening the tub?”

[answer] "Yes, it was difficult, it was too tight."

[question] "Do you have any idea why it is so tight?"

[answer] "...?"

☞ Another point of a more theoretical nature is the distinction for cognitions based on the tripartition suggested by Rasmussen (1990): ‘skill-based’, ‘rule-based’ and ‘knowledge-based’. This tripartition is included as it is often a reasonable method of classifying findings from empirical research. This applies especially to the tendency people have to stay on the lowest possible cognitive level, preferably ‘skill-based’ in the use automatisms, and only on a ‘knowledge-based’ level if nothing else works. Compare Kirlik (1994), who states that “cognitively intensive methods for action selection are used only as a last resort, ...” (p.83), and who typifies cognition as a ‘gun for hire’. As an example see the way pictograms on vacuum cleaner heads are ignored in figure 1.4

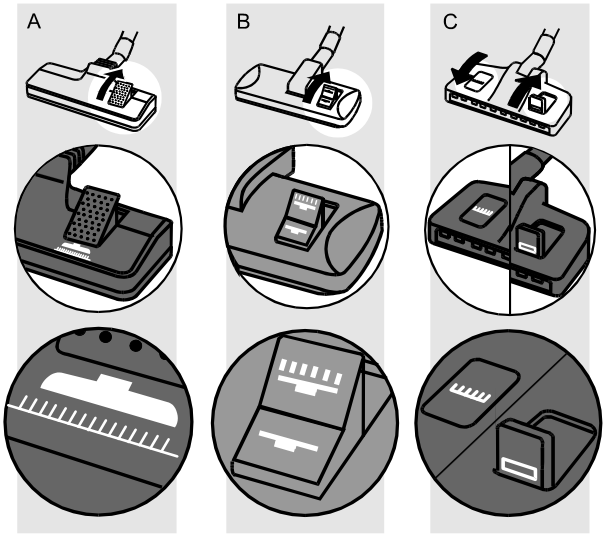


Figure 1.4 Pictograms on vacuum cleaner heads (from thesis, Loopik et al., 1994, see Kanis 1998)

The IIIIIII on cleaner heads A and B means ‘rug’ (deep-pile carpet), on cleaner head C it means ‘brush’, which is what all the respondents think it means for each of the heads. The pictogram with the IIIIIII on A means ‘brush in’. Those who think that it means ‘brush out’ can, for the pictogram on B with just a cleaner head, logically only interpret this as ‘brush in’. This is wrong: in the case of B if you see a cleaner head only this means you don’t see any carpet and that the cleaner head in question therefore works on smooth floors, in other words, ‘brush out’. In the case of cleaner head C the respondents benefited little from a correct interpretation: when using the vacuum cleaner you are standing behind the cleaner head and can only see the pictogram that is not relevant, as it is folded down, in other words ‘out’.

On the basis of such pictograms users won’t bother to try and figure out if the brush is ‘in’ or ‘out’; people will look, feel with their hand (not very hygienic) or notice from the resistance when vacuuming.

1.3 Human characteristics and qualities

Of course there is more to a user than the representation in figure 1.2, which depicts only use activities. In figure 1.5 the user is further defined on the basis of human characteristics and qualities: experience, current physical or mental state, and characteristics and capabilities.

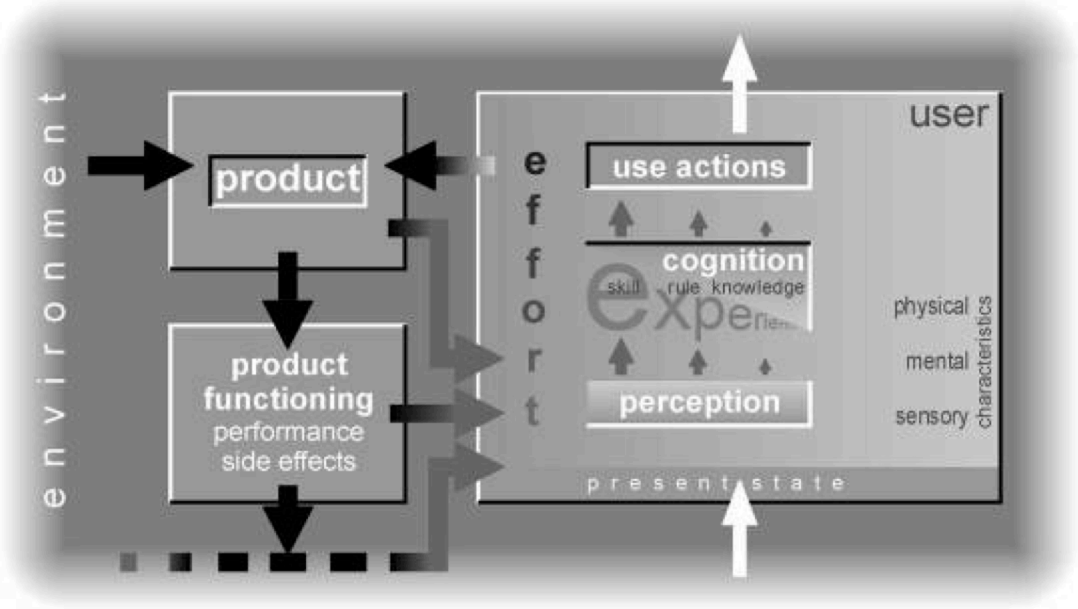


Figure 1.5 *The functioning of a consumer product (for earlier version see Kanis, 1998)*

➡ *Product performance (protection, support, replacement, expansion of human activities) and side effects (vibration, noise) as they are determined in a technical-physical sense by the concomitance of a product with use actions in a certain environment (see figure 1.1).*

➡ *The role of use activities-related perceptions: cognitions and use actions, including the effort involved in a product's performance; the lowest possible level of effort as default for perceptions/cognitions, in other words, preferably at 'skill-based' level (Rasmussen), compare how the size of the vertical arrows between 'perception' and 'use actions' increases to the left; experience as situated knowledge of users on the basis of association and recognition, and also as a psycho-locomotory condition, for instance typing skills.*

⇨ *Possible mutual influencing between the user and the performance of a product via the environment; 'present state' varying from physical condition (wet hands, fingers bandaged), sensory condition (otherwise always spectacles on) to moods of users (in a hurry, irritated, in a good mood or having a bad day).*

.

1.3.1 Experience

☞ For 'experience we can, in a use context, first consider experience as situated knowledge; in other words: knowledge about what and how, applied by users as a result of association and recognition in a certain use situation. This concerns a kind of similarity, but purely and solely in the eye of the user. What an outsider, like the designer in question, considers similar products/product components must be confirmed in practice: users may not 'buy' such an

intended/designed similarity. Incidentally, knowledge that is activated as experience may not be easy to apply in a certain context. For instance, most respondents in the creamer tub study get it wrong (see caption figure 1.3), fixated as they are on their experience that, one way or the other, creamer tubs always have to be opened completely.
Apart from situated knowledge, experience also manifests itself as a psycho-locomotory condition; consider, for instance, practical skills such as climbing/descending the stairs, riding a bicycle, using a keyboard. The effect of experience on the actions of users is, to a (large?) extent, always based on concrete, provable or demonstrable things – experience in general terms, like the way we 'occasionally use a computer', generally proves to mean nothing when it comes to how someone uses a 'similar' product. We will discuss this in more detail in the following chapters.

Figure 1.5 reflects experience as the counterpart to cognition: at a 'skill-based' level the source *par excellence* for how people do things, at a 'knowledge-based' level still relevant; consider, for instance, the ability to use certain knowledge in a certain use situation.

1.3.2 Current physical, mental state

☞ The second new term in figure 1.5, 'present state', refers to the physical or mental state of a user at the time of use: (physical) bandaged fingers, wet hands, (sensory) normally always wears glasses, (mood) in a hurry, irritated, enjoys using a product.
This is quite a varied collection of factors as, after all, 'present state' refers to the state of the various situations a person can be in. A practical example is the accident with a fitness stick referred to in the study by Weegels (1996). A mother wanted to show her children how dangerous this stick is. When demonstrating the stick it slipped out of her hand and hit her in the chin. According to Weegels' analysis the fact that her hand was wet may well have been a decisive factor.
Note that in figure 1.5 the current state of users is deemed to be partially determined by the influence of the environment (see bottom white arrow in figure 1.5). In this depiction of the process the functioning of a product may contribute, directly and/or via the environment, to the fact that the user is in a certain mood.

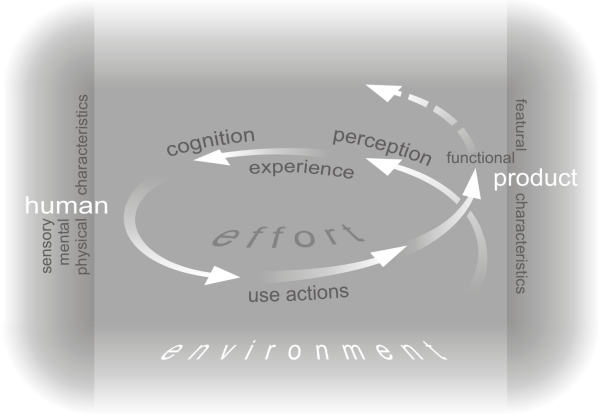
In other words, a factor like being irritated can play a role in user-product interactions as an external factor, for instance being irritated with people around you, but it can also be the result of use problems in the interaction itself.

1.3.3 Human characteristics

Figure 1.5 distinguishes between sensory, mental and physical characteristics in other words characteristics and capabilities of people; this refers to general qualities that are more or less constant qualities over an extended period of time, whether you are a user or not. Consider, for

instance, such basic characteristics as body weight and dimensions of (parts of) the body, such as length or girth, and also capabilities in terms of physical strength, joint flexibility, memory, strength of vision and hearing limitations, and also combinations of capabilities that manifest themselves in hand-to-eye coordination and reaction times. In general terms these are characteristics of people that have traditionally been the subject of studies in anthropometry, biomechanics, psycho-physics etc.

As a rule human characteristics, and more particularly differing capabilities, impose limits on the use activities of people. After all, the distinction a person can make depends on his or her vision, hearing sensitivity etc. Likewise, the speed of someone's reaction is linked to his or her ability to respond, or the strength someone generates with physical capabilities. In the next chapter human characteristics are discussed in greater detail as a possible source of explanation for what users do not do and also, with respect to their limited expressiveness, for what users *do* do.



Use as a succession in the context of use activities (... → perception → cognition → action → perception → ..., including the required effort) within parameters in terms of human qualities and product characteristics.

1.4. Interaction

Using the concepts of figure 1.5, the use of a functioning product can be described as a continuous interaction in the repetitive cycle whereby figure 1.5 is reduced to the core of what it is about, at least in the first instance.

1.4.1 Interaction

What does 'interaction' imply? This concept includes interaction as, for instance, between people who are talking to each other. In fact the term interaction as it is used here originates from the study into the (everyday) linguistic usage of people (for instance conversation analysis and in ethnomethodology). Can insights into the interaction between people be relevant for how users 'interact' with products? 'No' is the obvious answer, or at least not just like that.

Because even if you describe the functioning of (future) products as intelligent, or can see emotion-sensitive software on the horizon, all this cannot hide the fact that designed technical artefacts will, for now, continue to respond to use actions in a pre-programmed manner, no matter how 'human' the functioning of these artefacts appears to be (e.g. in the case of verbal instructions) or is described (compare the aforementioned robot). After all, everyday speech is quite different from (spoken) use actions. In other words, there is no fundamental equality in this case. This is why, seen in this light, the term interaction that originates from the study into the use of language is – to put it mildly – a distorted representation as a characterisation of the use of products by people.

However convincing this conclusion, based on observed (similarities and) differences between linguistic usage and use activities may sound, it is not the whole story. What do users themselves think? People will definitely blame themselves and not the hammer when they hit their thumb. But then again, often it is 'that stupid computer' that gets the blame when all kinds of strange things happen on the screen. Not surprising for something that has a 'memory' and is able to ask you all sorts of questions, but still… In short, from the user's perspective linguistic usage and use activities could sometimes be somewhat comparable. This may serve as a justification for the expression 'user-product interaction'. Incidentally, that still doesn't tell us what insights into linguistic usage could (in some cases) mean for insight into use activities. Before we discuss this a bit further (see next paragraph), let's consider another possibility for making both poles of the interaction in figure 1.5 – user and product – somewhat more matched, in other words, more equal from an interaction point of view.

This alternative way goes in the opposite direction from making products more human from a user perspective, as we discussed above, namely by what is called reificating the user, meaning 'turning the user into a thing' (from the Latin 'res'=thing, item). Once again, think of the computer as a source of metaphors for what happens in the calculation process in the human brain, the medium for the information process. An extreme representation in terms of this metaphor reduces the user to a kind a puppet, a conduit for 'input', 'throughput' and 'output' via sequentially organised procedures and behaviour programmes captured in internal representations. Compare, in this context, the so-called MPI model in the first year learning material (Dirken, 1997). Reification can quickly deteriorate into a behaviouristic representation of the actions of users, in other words a stimulus-response mechanism with a 'black box' in between; consider Pavlov's dogs.

Similar to an externally imposed 'humanised' depiction of products, turning users into more or less pre-programmed executors of internal behaviour programmes appears to be difficult to reconcile with observed use activities as described in this and subsequent chapters.

Incidentally, to what extent were we successful in not reificating the ‘user’ in the graphical depiction of the functioning of a product in use (figure 1.5)? Look at the different terms, linked by arrows or otherwise – it all looks, at the very least, reductionistic. But it is impossible to get away from making distinctions – holism doesn’t get you anywhere.

In addition to a range of items not being named or specified in figure 1.5 (see the end of this chapter), in the graphical depiction of the functioning of a product in use parallel processes are indicated in the involvement of users, something that makes this involvement much less rigid than a serial depiction, compare the MPI model (first year), in which you always end up in behaviour programmes via decision processes. Perhaps some ideas for a less mechanistical depiction may present themselves.

1.4.2 Linguistic usage and product use activities

As far as we know, in earlier research linguistic usage between people and use activities of people with regard to products have not or hardly been compared. But linguistic usage is very much a factor in another way, namely as a tool in the study of product use. In particular, this relates to the internal processes of users. This makes a user not only the most readily available source of information but also the only one, with language as the method *par excellence* to access this information. In terms of figure 1.5 this relates to:

- the user's observations in a certain context (‘perception’),
- the thought processes of the user (‘cognition’),
- the experience which the user deems applicable to the situation in question (experience’),
- the required effort (‘effort’), and
- the (emotional) state that may be relevant for the conditions in question (‘present state ’).

Later, in chapter 2, the fact that these types of issues must be raised as unemphatically as possible, for instance in an open interview, comes up. Insofar as people are able to talk about ‘what they felt’ – unlike in the case of natural automatisms this is not always possible, see also chapter 2 – such an open interview will soon start to resemble a normal conversation. This means you can’t get away from the question how people use language to assign meanings to the world around them, to their own sensations, in short: how language works.

In language, is only a reality being reflected that is in fact ‘objective’, retroactively ‘labelled’ with adequate terms based on the perceptions and experiences of people? This is a view that, in a number of gradations, is accepted on a broad (cognitive) scale. It provides – at least in the mind – a reasonably structured, understandable world, with people who (can) talk properly and say what they mean, something that is not in fact the case at all, according to language research. The way people talk to each other is frayed, apparently contradictory and inconsistent. These kinds of findings match the view that language is not a passive documentation system at all but (in fact) a means to create ‘reality’. This somewhat less purchase-providing view, that meanings are

(partially) given shape in the linguistic usage of people, harks back to the work of the Russian psychologist Vygotsky (1934). The 'reality' is not set in stone but is recreated each time in differing contexts, and is therefore 'discursive' as this is sometimes called in research into linguistic usage (originates from 'discours', see Edwards, 1997). Of course it does not pay to exaggerate this. Someone running into a lamp post will notice that there is little discursive about it. In the research into linguistic usage the emphasis is on how people talk to each other and what comes to the surface in their linguistic usage, for instance how they negotiate about each other's (assumed) interests.

It is not an immediately obvious course of action to make a discursive analysis of an open interview between an interviewer and a respondent, for instance about tracing the experience that
! was relevant to the respondent under the given circumstances (see §1.3.2). Still, this doesn't say it all. After all, respondents need not have a clearly defined idea about this; they may, in addition, randomly bring something up, be trying to look good to the interviewer (socially desirable), or afraid to be considered stupid otherwise.

The study of linguistic usage, in combination with other readily available evidence such as video footage of use actions, can provide insights about this, see the following example from the study by Rooden (1998) into the use of a sphygmometer.

[question by researcher]	'Are you normally speaking left-handed or right-handed?'
[answer by test subject]	'I am right-handed and all my injections have been in my right arm. And I row on the starboard side, I am very consistent.'
Comments:	Even though this participant indicates that he is right-handed, he still pumps with his left hand during the test. He piles up a number of things that are related to being right-handed in the answer to this question. Is he trying to come across as convincing?

This is as far as we will go on this tangent. Later in this chapter, and in chapter 2, we will come back to linguistic usage and (the giving of) meaning.

1.5 Usability

With the representation in figure 1.5 in mind we can now give a further description of the concept
☞ of 'usability':

a An adequate product performance ('product performance' in figure 1.5), this means that the design functions 'effectively' in the protection, support and replacement of human activities, and is also effective in the expansion of human capabilities.

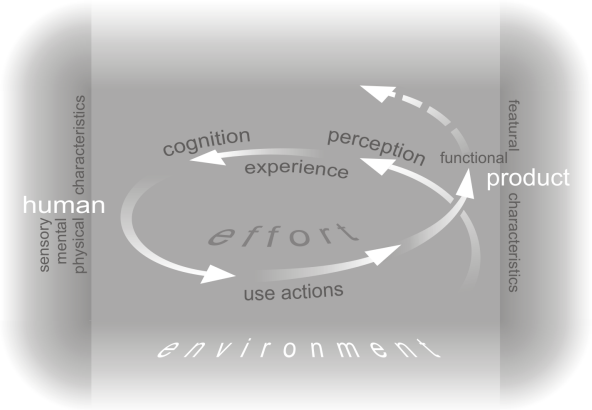
b Minimum side effects in the functioning such as vibration, noise, (the risk of) accidents.

c A desired physical and mental effort ('effort' in figure 1.5) associated with use activities; as a rule this effort is preferably as minimal as possible, whereby a distinction can also be made based on:

☞ • guessability at the first use (or after long period of non-use), and

☞ • learnability of use activities, to get the maximum benefits from a product for the least effort (Jordan et al., 1991)

☞ A combination of a and c gives the term 'efficiency' as the relation between the product performance of the design and the effort required on the part of the user to achieve this performance. Please note that the term 'usability' is – not surprisingly – only given substance in terms of the repetitive cycle.



☞ The list of general criteria for the usability of a design therefore looks as follows:

- the level of the product performance, or the effectiveness,
- the extent of the side effects, and
- the required effort, which can also be expressed as efficiency (in other words, effectiveness in relation to effort).

Satisfaction, often referred to the same breath with usability, is not included in this list. Why is that?

1.5.1 Usability and satisfaction

☞ The English term for the Dutch word 'bruikbaarheid' is 'usability'. International agreements were

☞ reached about the content of this term, see ISO 9241 (1994). According to this international standard 'usability' is, in addition to 'effectiveness' and 'efficiency', is also dependent upon 'satisfaction'. The reason for user dissatisfaction is often effectiveness that is experienced as being below par, a number of side effects and/or the experienced effort, compare Jordan (1998). In many cases the reason for such dissatisfaction can be summarised as experienced discomfort.

From this viewpoint the ‘dissatisfaction’ experience does, in the first place, say something about individual references of people for classing something as uncomfortable, bothersome, disappointing etc. When designing based on the usability of products the personal aspect of such an experience does, in itself, add little of value: the effectiveness, side effects and required effort were either already good, or – in contrast – not.

So is it true that the smooth functioning of a product, that does exactly what you expect it to do, without any problems and for an effort that is hardly noticeable gives users satisfaction with regard to the usability? Given the definitions there is little sense in doubting this, certainly when it comes to using a design for the first time when you have been used to something worse. Except
! that this new usability tends to become the standard – or ‘default’ – sooner rather than later. This touches on the core of what has been referred to as the ‘snag’ of ergonomics: if ergonomical insights are applied properly in a design you don’t notice them. To put it differently: the absence of negative effects as the highest achievable goal. Certainly, this is phrasing it extremely. But compared to something as common as ‘content thanks to the Nike feeling’, lasting feelings of satisfaction in users will often originate from aspects you cannot clearly call ‘good’ or ‘bad’ in terms of usability. Consider the fascination with a certain shape, the sound of a closing car door, the terrific feel of a control switch via the perceived effect. All these emotions are extremely relevant to design. But this is no reason to attach emotions to the term usability as such. In short, we will stick to effectiveness, the occurrence of side effects and efficiency as criteria for usability.

In the proposed concept delineation nothing has yet been documented about usable according to whom.

1.5.2 Usable according to whom?

So can you say that the user is the measuring standard for all things? A positive answer to this
! question is a good aspiration. Except, there are so many users. What one considers effective is merely so-so for the other, for instance depending on what you are used to. Equally, what is an acceptable nuisance for one may be unacceptable for someone else, compare figure 1.6. And the same goes for the required effort, see figure 1.7. ‘The user’ has many viewpoints.

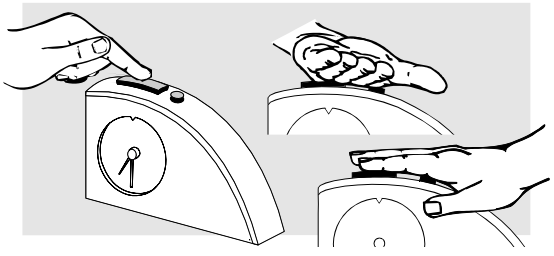


Figure 1.6 *Experiences with a clock radio (Voûte et al., 1993 (thesis))*
First of all this picture illustrates the problem of the simultaneous operation of both controls on the top: the large button is for the alarmclock, the small one for the radio.

Those making the effort to aim properly (respondent on the left) avoid having to get up afterward because the radio isn't doing what you want it to do after you have hit it to turn off the alarm (both respondents on the right). Another nuisance factor to which people proved to react differently in this study is the clockface illumination. This bothered three of the eight respondents, even when set to the lowest level. Measures of these respondents: turning the alarmclock around or putting some books in front of it, as a result of which you can no longer see at a glance what time it is. The other five respondents had no problems in this respect.

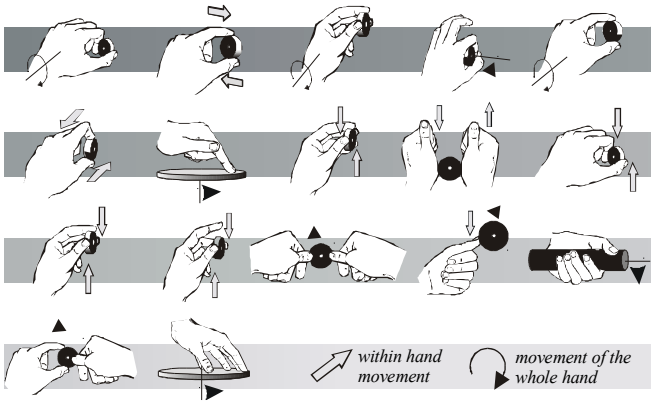


Figure 1.7 *Operation methods of round turning knobs, observed in two field studies (Schoorlemmer and Kanis, 1992; Kanis, 1993).*
People with and without physical disorders were involved in these field studies. It was noted that people with arthritis (a common type of rheumatism), who cannot operate the knob without using two hands and who really always experience pain, still class the operation as problem-free. Firstly this is a matter of individual references, depending on personal circumstances and experience. The other thing that may play a role are differences in the interpretation of words like 'exertion', 'effort' and 'problems'.

Respondents appear to have a tendency to express more satisfaction than you would expect, given their fumbling and sighs when using a product. In short, giving 'the user' the last word may get you off on the wrong foot. Sometimes the user does not even have this last word for a good reason, see the example in figure 1.8.

! Even if a user himself has no complaints – not knowing any better -, thanks to observation of how users go about things, for whatever reason, the designer can most certainly end up with a problem. In the case of figure 1.8 this is how to make the vacuum cleaner 'explain' more clearly, 'radiate', what it can do and how these possibilities can be realised in its use.

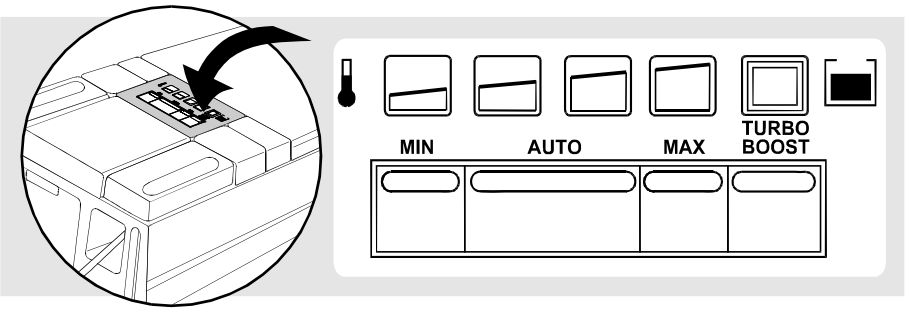


Figure 1.8 *The electronic suction control on a vacuum cleaner (Loopik et al., 1994 (thesis))*
The term AUTO means that the suction is automatically adjusted depending on the type of floor covering. A user who is shown to operate the vacuum cleaner as if AUTO were the correct setting for vacuuming the car need not necessarily get the impression that it could be better, or that vacuuming takes a relatively large amount of effort. This may change as soon as the user becomes aware of the functionality available in the design and also what operation is needed, for instance because for some reason he finally consults the manual, or because an outsider points it out to him.

! In another way users who are – in their mind – not experiencing problems still cause designers to scratch their head, see the example in figure 1.9.

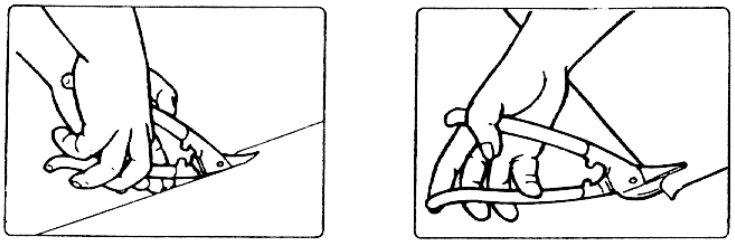


Figure 1.9 *The use of snips*
Left: how designers thought a pair of snips they designed would be used. Their design was based on this use method. Right: how the vast majority of users use the snips, according to use observations. This actual way of using the snips proves to be more efficient (less tiring, better view of the work); Sperling et al., 1993.

! With figure 1.9 we must note that it is almost certain that, in this case, the designers missed some opportunities. In other words: bearing in mind the knowledge demonstrated by figure 1.9 a (still) better design appears feasible. Or to put it differently, only after use observations can designers in these examples conclude that they do not have design problems when users indicate that they themselves do not experience problems using the product. The reverse does not happen: a use problem, in terms of reduced usability, cannot be ‘patched up’ in any way, i.e. when it comes to usability problems experienced by users, the users definitely have the last word.

Whether a usability problem also becomes a design problem depends – apart from available time and resources – on the following considerations that must be partially weighed against each other:

!
- the seriousness: for instance is there a chance of personal injury,
- the scope: how many users will be affected by the problem, and
- the permanence: compare ‘teething’ problems that disappear with a little more practice versus persistent problems that sometimes don’t show up until after longer use, see the example in figure 1.10.

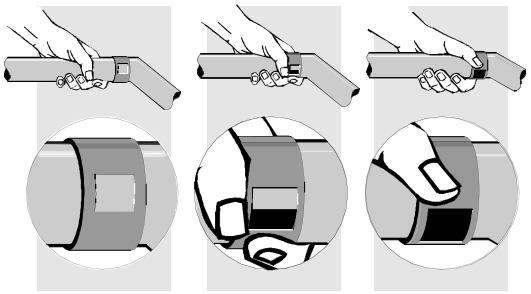


Figure 1.10 *The mechanical suction control on a vacuum cleaner handle (Loopik et al., 1994 (thesis)) In the way of holding the handle shown in the centre and on the right the mechanical suction control is inadvertently opened. The result is reduced suction. Because the opening is on the bottom of the handle respondents used changes in the sound to see whether the control was open again. This problem emerged in the second observation, after the vacuum cleaner had been used for a week, and not during the first time the vacuum cleaner was used, immediately after it had been unpacked.*

1.6 What concepts are missing in the representation of the user-product interaction?

The question how complete figure 1.5 is relates to both the concepts used to structure the depiction and the links between these concepts. The latter aspect, the relations between concepts, is discussed further in the next chapter.

1.6.1 Personality characteristics

With regard to the terms in figure 1.5 we have already referred to the colourful collection of more or less temporary, coincidental characteristics of people, such as ‘current state’, from wet hands to being in love, in a manner of speaking. Should we not also add the personal traits of people that are considered more permanent, the so-called personality types?

A well-known typology for these characteristics is the one that is based on combinations of scores for different 'dimensions', such as <introvert – extrovert> and <rational – emotional> (developed by Myers and Briggs, who base their theories on the Swiss psychologist Jung). A typology that is currently getting a lot of attention is the so-called 'Five-Factor Personality Inventory', resulting from answers to questions that are scored for the categories 'extrovertedness', 'mildness', 'orderliness', 'emotional stability' and 'autonomy' (Hofstee, 1999). Attempts to isolate these typologies as archetypes from words in dictionaries for various (western) language areas do not go unchallenged. Different cultures, also over time, need not result in a single (same) typology, something that applies equally to emotions (Edwards, 1997). Furthermore, the use of language is once again the issue here (see §1.4.2), which, of course, is something quite different than starting with words that have been extracted from dictionaries. With (real) linguistic usage as a reference, ideas like the 'Big Five' can never have fixed meanings. The range of issues raised by users is much too varied, changeable and contextual for this. A reduction of this diversity into (assumed) 'archetypes' or 'prototypes' does, as a rule, result in a limited number of categories that are just as general and meaningless when applied to the individual experiences of people.

Incidentally, the authoritativeness of a categorisation like the 'Big Five' appears to be very remote anyway for the actual use of products. For what they are worth, this does after all relate to typically psychological ideas, in other words inherent (depth) characteristics of people, not based on how people interact with a product.

1.6.2 'Decision-making', 'problem-solving', mental models

What is also missing in figure 1.5 is a term like 'decision-making', a concept that is referred to in the aforementioned MPI model (Dirken, 1997). Making decisions as part of the use of a product does, like the expression 'problem solving' in this context, remind us of cognitively rationally operating users, who choose the best solution for them on the basis of a systematic analysis of alternatives. Perhaps such a depiction is useful in exceptional cases where the use occurs on a purely 'knowledge-based' level. But all the empirical material that has served as input for this reader (observations of respondents using familiar and unfamiliar products) does, at the most, provide indirect indications that there may be an occasional rationally operating user, as a problem-solver.

Anyway, how far do you go in assuming all kinds of mechanisms in the heads of people, such as designing alternatives, maintaining associated hypotheses, making considerations between them. Often there are more simple, direct and obvious explanations, namely in what people bring up themselves. The problem with depictions of what happens in people's heads is always that you can not or hardly get at it, that you can never prove this is the way it really happens. This is also exactly the speculative aspect of so-called mental models of people; for an overview article, also with regard to all the confusion, see Wilson & Rutherford (1989). Mental models are abstractions, in whatever philosophy or system, of the knowledge of people, their expectations, their

experience. In fact you could regard a mental model as a disguised ‘homunculus’ (Gelderblom, 2001), a little man in our head that regulates everything insofar as we can’t reach it. An important characteristic of this little man is that it, too, has an homunculus in its head, that regulates everything insofar as we can’t reach it with our first homunculus. And an important characteristic of the second homunculus, in the first homunculus, is that it, too ...

And so we go on mindlessly pushing questions ahead of us, in what is sometimes called an ‘infinite regress’. Furthermore, if you have to keep going back to terms like ‘knowledge’, ‘expectation’ and ‘experience’ to flesh out the abstract concept of ‘mental model’, this entire abstract concept is useless. Worse still, sometimes it is, apart from a source of misunderstandings, also meaningless because of its suggested comprehensiveness. This substantive emptiness is most strongly expressed in the inability to falsify pronouncements about someone’s ‘mental model’, in other words: to determine, on the basis of empirical observations, that you do not have mental model this but mental model that. On this issue see also Gelderblom, 2001.

1.7 ‘Situatedness’

☞ In the depiction of the user-product interaction in figure 1.5 references to mental processes as referred to in the previous paragraphs are missing. This omission is intended to show that such processes are considered of secondary importance in view of the findings of empirical studies. But is there an alternative? And does the depiction in figure 1.5 leave room for this? Yes, an alternative that may be suitable to a certain extent could be the characterisation of the use as ‘situated’ (Suchman, 1987). This means the use is interpreted as being strongly context-dependent (compare against the previous paragraphs about experience as situated knowledge). Furthermore, the concept of ‘situatedness’ reflects more than anything that there are few points of departure for the theoretically substantiated prediction of use activities. This gives the concept a predominantly pragmatic overtone, in the consequences for the structure of research into person-product interaction (see chapter 2).

Insofar as product use, in its diversity, reminds us (yet again) of linguistic usage, a parallel between the two is not entirely coincidental in the often-quoted research by Suchman into the use of a photocopier. This is because in this study Suchman asked her test subjects to work in pairs, so generating a lot of language use. Another example of this approach is the research conducted by Frolich, Drew and Monk (1994), into what people do when something goes wrong when using a software programme. Suchman emphasises the improvising character of users in their actions and activities. This does not preclude the possibility of simultaneously typifying users as rule appliers (Garfinkel, 1967), consider use activities at skill and rule-based level (Rasmussen). Also compare the fixation of people resulting from a one-sided experience referred to in §1.3.1, something that is further discussed in subsequent chapters. Considering people as rule appliers

is not the same as ‘pattern followers’ (Reason, 1990). This latter typification reveals the standpoint of an observing outsider, as if individuals would somehow, in their behaviour, adjust to one another, as congeners with all kinds of commonalities, such as reasoning methods and ‘behaviour patterns’.

An example of a more general trend in the application of rules by individual users is the lowest possible cognitive level as the default (see figure 1.5). In the next chapter, in which the (un)predictability of use is the focal point, more of these general trends will be discussed, in which the lack of the terms ‘plan’ or ‘intention of the user’ in figure 1.5 is also discussed. Had you already noticed that these terms were missing?

2 Conducting use research

☞ Like any research, use research has an objective: the *why* of your observations, in this case in a design context. This *why* is different from what you will observe – with regard to the study questions. The design phase in which use observations are performed does, in general, have consequences for the set-up of the research, in particular whether or not simulation is required. In this context the issue of how to realise the greatest possible authoritativeness of observations is really relevant in every phase of the research. In other words: how to prevent findings being distorted, or biased. With these points for attention in mind – the type of research being dependent on the design phase and the prevention of distorted findings – this chapter deals with the structuring of use research, from rendering presuppositions more explicit to translating findings into design solutions.

2.1 Objective of use research in a design context

☞ Use observations must provide an insight into the what, how and why of product use, in other words, in the repeating cycle of the use, involving the possible role of human states such as experience and general characteristics (see figure 1.5). The importance of such insights is demonstrated once more by table 2.1: what could go wrong, or at least different from what was intended in a design with regard to perceptions, cognitions and use actions.

material and functional product characteristics

not observed, reason:	*observed; intended meaning not understood, reason:*	*observed; understood as intended in design; not acted as expected, reason:*
• not directly observable, e.g. on/off switch on back of videorecorder • missed, inconspicuous, escaped notice, such as icons on a screen • not knowing that there is something to observe, not familiar with functionality such as 'mute' on a remote control • not being able to 'place' reference, failing to recognise, familiar with functionality such as two-digits channels on remote control	➢ too much effort and obvious alternatives, such as icons on vacuum cleaner attachments, see figure 2.4. ➢ misunderstood, such as the term AUTO as 'car' in figure 2.8 ➢ allocated meaning not envisaged by design such as taking **seams** to be openings (Loopik et al., 1994)	⇒ own alternative is preferred (more effective, more efficient, …) ⇒ mistake, as in **wrong** test ⇒ delay, user thinks he can do something else first ⇒ ……

Table 2.1 Use activities different than (presumably) intended in the design for a variety of reasons

As we commented before (in different terms) when it comes to what a designer has to do with regard to (redesign) interventions it makes quite a difference whether a use problem is caused because designed use cues were overlooked, not recognised as such, not or wrongly understood, whether there is a question of false use cues, etc.
The question, then, is what the possibilities are for incorporating such insights, preferably early in the design process (a new design that is, at that time, still in *statu nascendi*), with the advantage that at that stage all kinds of things can still be changed at little cost, and the disadvantage that there is not yet a tangible, material design that can actually be used to actively try different options.

2.1.1 Use observations in different phases of the design process

The depiction of matters in figure 1.5 relates to a functioning product. But this is therefore not the only form of use research that is relevant within a design context, see figure 2.1.

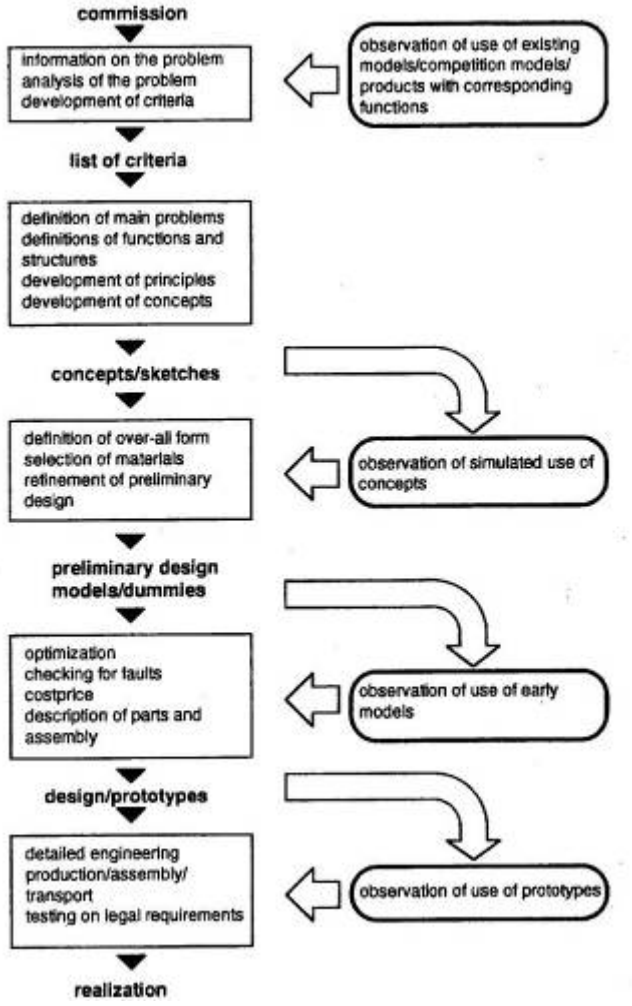


Figure 2.1 Research with users in the design process (Marinissen, 1993).

Left column the depiction of the design process, in the main outlines derived from Roozenburg and Eekels (1995), on right possible input from van use observations, at every ⇦ .

Figure 2.1 shows four successive observation moments:

a observation of the use of an existing product,
b simulation of use on the basis of concepts, such as drawings,
c as in b, with primitive models, in other words, not of or only partially functioning, and
d use of a prototype, in other words, functioning to an important extent (but not 100%).

In case *a* it usually concerns products of unknown design origin and you can not refer to designed use cues, at most assumed use cues. In cases *b*, *c* and *d* designed use cues are known, insofar as the designers involved feel they can make pronouncements about them.

We always assume that in cases *b*, *c* and *d* all design considerations relating to material and functional product characteristics of a model or prototype are known in the realisation of the simulation or observation of the use – after all, ideally the use researcher is (one of) the designer(s) involved, see more about this in §2.7.2, *Who conducts the research?*

Another characteristic difference between variant *a* in figure 2.1 and the rest is that the full functioning in *b*, *c* and *d* always involves simulation to a greater or lesser extent. Simulation is not possible without express instructions to test subjects about material or functional characteristics of the product or the product components that are still missing and that are essential for the simulated functioning. This has consequences for the authoritativeness of findings, as the more pointed a research context, the more artificial elements can have an interfering effect compared to a natural ‘situatedness’ of the use. The result is that findings can easily be biased, something that touches a sensitive spot in any kind of research: the occurrence of distortion.

Before we talk about this in the next paragraph, first a brief note with figure 2.1:
In the description of the design process by Roozenburg and Eekels (1995) the term ‘analysis’ is used for the phase in the process in which all the information that has been gathered up to that point is analysed, in other words: broken down into its components, assessed and evaluated. In the Roozenburg-Eekels description of the designs the actual conducting of research is not mentioned in so many words. If any research is in fact conducted it can be assumed ‘hidden away’ in the information phase. Figure 2.1 shows this is much too limited for conducting use research which, after all, can be done at many different times in a design process. In the Roozenburg-Eekels’ terminology framework this can then be given a twist, via the cyclical character of the design process.

Truly satisfactory this is not. The question arises whether the cumbersome fitting in of research is the repercussion of historical practices, in which design and research had little or nothing to do with each other.

Both activities have been and continue to be regarded as being diametrically opposed, whereby research is presented as being hypothesis-evaluating (compare Eekels and Roozenburg 1991).

☞ This is a type of research in which a hypothesis is evaluated by means of the controlled manipulation of research conditions, thus putting a theory on the 'torture rack'. In this context the term 'hypothesis' refers to a statement/assertion on the basis of a certain theory, in other words demonstrably deduced from that theory. Hypothesis evaluation is a type of research that textbooks often fall back on, compare De Groot in his textbook on research methodology (1960, re-published some years ago).

! But it is nowhere near the only occurring type of research used. Especially in the area of person-product interaction, where more comprehensive theories are not or hardly available, empirical research is, to an important extent, explorative in nature. Kirk and Miller (1986) point out that the most drastic discoveries were specifically not made in hypothetical-deductive research: "In testing a hypothesis, you know what you're going to discover." According to the aforementioned authors this form of research is "aimed at preventing discovery" (P. 15). In the IO course you are likely to come across the requirement that hypotheses must be formulated before the research can start. Ask yourself whether, in such a case, this is not simply 'fussing' about slapdash work, obvious ideas or whether suggestions are being blown up into hypotheses. Because why, as brave people, should you comply with that, slavishly following the rules.

With a somewhat broader understanding of the various forms of research it is possible, according to figure 2.1, to make a productive contribution to research about design in a relatively easy manner.

2.1.2 Distortion/abnormalities

☞ The most important threat to the authoritativeness of research findings is the occurrence of distortion or abnormalities, in two respects:

I you register or observe something other than you think you do;

II you wrongly generalise.

(ad *I*) Some people who claimed not to have problems opening a new jam jar admitted, when asked further, always get someone else to do this for them (Steenbekkers, 1998). Did these people give a wrong answer? Now, they answered a different question than the interviewer thought he was asking. Furthermore, taken as a whole, the effect is that a more rose-coloured, in other words, somewhat distorted picture is created of the problems associated with opening jam jars.

(ad II) In the study into the use of turning knobs referred to earlier (see figure 1.7) an attempt was made to predict practical problems that people have with the required use of force, on the basis of measuring the actual force exerted with the hand under standardised circumstances (a prescribed posture and two prescribed ways of manipulation), see Kanis, 1993. First of all, table 2.2 shows that the prediction is correct in only just over half of all cases (all cases on the diagonal). This corresponds with the findings of Steenbekkers with respect to the limited authoritativeness of general human capabilities during the course of the use. In addition table 2.2 shows that the R-ratio overestimates problems (four times as many cases in the top right as in the bottom left). One of the more or less obvious explanations for this is the fact that people can, in practice, exert more force thanks to their own choice of posture and different than those prescribed in the force measurements. The relevant issue here is that table 2.2 shows that you cannot simply extend findings from standardised, controlled circumstances to 'the' practice, that you will quickly be faced with distortion, also referred to as bias.

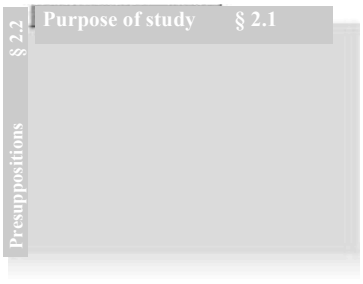
opinion of test subjects	*R<0*	*0≤R≤1*	*R>1*
problem-free	42	18	14
with effort	11	26	33
impossible	2	3	22

Table 2.2 Required force of use Opinion of test subjects regarding the force required to operate a knob, Fop, versus prediction on the basis of the ratio R=(Fop-Fcomf)/(Fmax-Fcomf), with Fmax the maximum force a respondent can exert, and Fcomf the comfortable force expended, in other words, without pain or extra effort (Kanis, 1993).

You must always be prepared for possible distortion, in observations themselves, and in generalisations of these observations. The prevention of distortion is a material characteristic of good research. Below a separate paragraph follows in which terminology that is commonly used in this context, such as 'validity' and 'reliability', is discussed.

2.2 Presuppositions

After this the treatment of the conduct of research will be documented in this chapter by means of a diagram that will be expanded bit by bit. In this diagram, of which we have discussed the research objective, we have now reached the presuppositions.



The diagram indicates which subjects have already been discussed, in which paragraphs, so that is easy to look back and find them. The sequence in which the subjects are discussed is, in large outlines, based on successive phases in the research. But do not consider this sequence as a strict

leitmotif that must be followed to the letter. As a rule all kinds of things happen at the same time, a lot of things are interrelated, later decisions can make it necessary to revisit earlier ones (compare the cyclical character of the design process).

Every study is based on a plethora of presuppositions, however hidden or obvious starting points often are or appear to be. Those who think they are looking from a clean slate, as if the spirit was a 'tabula rasa', would see nothing – data never speaks for itself.

An example. As a result of the kidnapping of Heijn (brother of our country's largest greengrocer) some years ago, the then Minister of Justice rang the alarm bell over the Netherlands as a refuge for the internationally branched out gangsterdom that, after all, had a hand in this kidnapping. However, research in 'that' direction came up empty. It was ultimately thanks to the attentiveness of two local police officers, as a columnist wrote, that the kidnapper and murderer (who was operating on his own) could be apprehended after he had used the marked ransom money to pay for something.

Presuppositions and the way in which they form the basis of study questions and structure, must be made more explicit before the research commences. Why? Because presupposition is the thought/reasoning as to why you think that something is the way it is or could be. Once the presuppositions have been formulated it is revealed, in so far as this is possible, why the study is looking at this particular aspect and not another, why it is researched this way and not the other way.

In this respect, what is wrong in the following example?
At night a man is wandering around in the light of a lamp post. A passer-by asks what he is doing.
"I am looking for my keys", is the answer.
"Strange place to lose them", responds the passer-by.
"That may be so", says the man who has lost his keys, "but at least here I can see where I am looking."

As mentioned before, presuppositions make it clear why, in a study, some points are questioned and some are not. For instance, the (in)conspicuousness of designed use cues is a possible point for attention, as a plausible presupposition that the material or functional directions for use in question are or are not being observed ('perception' in figure 1.5). In this context perceptions may be associated with specific experience ('experience' in figure 1.5). Generally speaking, experience, as situated knowledge on the basis of association and recognition by users, forms an important source of presuppositions for the assumed understanding of designed use cues, or for the unintended interpretation of cues ('cognition' in figure 1.5). In the same manner, incidentally, unforeseen use cues are, as a rule, derived from knowledge based on experience.

When thinking through the study structure making presuppositions more explicit behind possible material and functional directions for use will, in practice, soon result in the fact that a list of use cues is formulated. Two situations must be distinguished carefully: whether they are designed or assumed use cues.

☞ **Designed use cues**

In most cases the number of material and functional product characteristics actually designed as directions for use are limited. In the design practice it often does not happen, because, in addition to the future use, there are so many other design problems that require attention. Please note: designed use cues have nothing to do with all the wonderful things that can be 'thought up' afterward. However, a use cue *can* be co-designed as a matter of course, from a second (design) nature. The design story – which alternatives were considered with a view to the future use, what is (sometimes implicitly, therefore) rejected based on that viewpoint, or rather applied – that is what it is about. So no invention, simply to have some use cues, as if you have to. It frequently happens that, when setting up the use research with a design prototype, a design team has to conclude that no use cues were designed: ran out of time, glossed over it in the textbook, too many other things on the mind, or, what is also possible, deemed unnecessary because the designed prototype is deemed to speak sufficiently for itself with regard to its future use, without extra use cues. As a designer make sure you are clear about this, rather than providing 'retro-use cues' as a result of socially desirable behaviour: 'you ask (which is not the case at all), we deliver'; in this context see also §2.5.5, *Question & Answer*.

Assumed use cues

Practice teaches us that designating assumed/possible use cues to products of unknown design origin will quickly deteriorate into comprehensive invention: every colour difference, any seam, every curve could possibly have a deeper meaning. The result: dozens of use cues, presumably mostly seen as working independently/in isolation. Certainly, making designers indicate what they have put into their design in material or functional directions for use need not correspond with what designers think will work in the use practice as a use cue for a product of unknown designer origin. But compared to (mostly) no more than a handful of truly designed use cues in the design practice a number of several dozens of assumed use clues seems a little haphazard, and also noncommittal. This big difference would mainly seem to indicate that designers are good at inventing things, or that the term 'use cue' invites them to do so, or both. So how do we sensibly limit the number of issued use cues? Because, admittedly, the screen will soon be full of possible cues. Restrict yourself to material and functional product characteristics that may be considered essential (their absence would reasonably mean a considerable chance of use problems), and would also be notable, unmistakable extras instead of common in terms of cultural conventions.

All this illustrates once again that use cues are negotiable, will vary inter-individually in the meaning given to material or functional product characteristics, whereby tracing this meaning is

first of all discursive (compare §2.4.2), in other words it will need to be discussed. Another argument in favour of not coming up with the pretend certainty of rows and rows of use cues in so many principal uncertainties.

We commented on this before: whether it concerns designed or assumed use cues, often knowledge based on experience constitutes the assumed basis of the (non) functioning of material or functional directions for use. Another source of possible presuppositions is the characteristics and capabilities of the human organism, the 'characteristics' in figure 1.5. Consider anthropometrical characteristics, eyesight limitations, in the exertion of force. In general presuppositions, derived from such characteristics, will contain pronouncements about what cannot be done and why. The characteristics in question impose limits on use activities of people (see §1.4 at the end). Such preconditions prove to be not or hardly indicative of what happens with regard to use activities within these boundaries.

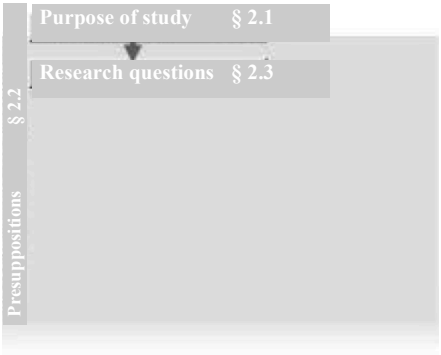
To reiterate: a presupposition is more than an expected or assumed outcome; namely it concerns also and especially the why of such a possible outcome. Therefore, do not confuse presuppositions with obvious predictions of what test subjects will do and not do that no one has any use for, if no explanations are given: for the latter you are conducting use research. So, in the preparation of a use study do not come up with such predictions as "test subjects will observe and understand the calibration next to the switch, as well as the red flashing LED, after which the product will be turned off." Exactly the same that applies to knowledge and experience applies to the possible role of human characteristics and limitations: if effects of differences therein are suspected, indicate why something is the way it is, and not otherwise.

An illustration. Let's say that results from an earlier study, or theoretical insights, give reason to assume that there could be a relationship between having a certain experience in the use of an existing product on the one hand, and use actions with a new design for the product on the other hand. Phrased this way, this presupposition is still hardly specified. The pronouncement that the use of the new design will result in fewer problems as users have less experience with the old product is already taking it further: there is clearly a presupposition behind it. It would go even further to predict that, for the intended effect, a use frequency of once a month is essential. Easy to test, such a prediction, but why, based on what? If there is no answer to this question such a prediction does not mean much more than thumbsucking, something that can easily give an entirely wrong twist to the study. Stay well away from that, and submit reasonable arguments for every presupposition (which in itself is permitted to be risky). This means: refer to existing theories, to published research, to other documented evidence, possibly own experiences as long as they can be concretised and preferably verified.

To conclude this paragraph the following. For the purpose of suggesting assumed material and functional directions for use, frugality must be the *leitmotiv*, rather than uncertainty – N.B. the

reason for the study – as a licence for inventing one use cue after the other. Reticence must also be the message when making all kinds of other presuppositions that are not directly related to the functioning of material or functional product characteristics as directions for use. Consider required effort depending on physical limitations, or the possible conditioning of use activities through other human characteristics such as sensory thresholds or memory functions. Not that there would never be sufficient reason to come up with the presuppositions in question. But in the same way that every advantage is said to have a disadvantage, continuously making up new presuppositions also has a flip side: the risk of a facet-view, with an eye only for all kinds of invented possibilities, this at the expense of a broad view, of being open to never assumed events. Don't forget that use research is first and foremost explorative. You must not 'use cue' such research to death or presuppose it to be useless before you've even started.

2.3 Research questions



☞ Research questions are sentences that finish with a question mark (Kuipers,1984). Research questions must be as concrete as possible, so they can be answered through actual research. Generalities of the type of 'how can we improve the world?' can be avoided by concretising your affairs without deteriorating into fragmentation such as asking ten or twenty questions about all kinds of particulars, about this detail and that, which causes you to lose sight of the bigger picture.

Because use research must answer questions about the 'what' and 'how' of use it is difficult to avoid a question like:

- What do people do and omit in the actual use of the model/product, in particular with what problems in terms of poor/inadequate usability?

And because use research must also give answers as to the 'why' of the use you will soon arrive at a question like:

- What role do material and functional product characteristics, designated as designed or assumed instructions for use, play in the occurrence of use problems? Where are these problems demonstrably the result of the lack of adequate use cues?

And another one in the same spirit:

- If there are no problems do designed or assumed material and functional instructions for use help things go well? If so, the next question may be whether this is not looking at things too reductionistically.

Other possible questions are:

- Do test subjects perhaps detect use cues in characteristics or in the functioning of the model/product that were not expected in terms of designed or assumed material or functional instructions for use?
- Do differences in certain experiences of test subjects play a role in what goes wrong and what goes right?
- Is the occurrence or absence of use problems associated with differences in human qualities, capabilities?
- To what extent are observed use problems of a temporary nature, a matter of getting used to a new product? And, after longer use, do stubborn problems perhaps emerge that did not manifest themselves in the first encounter with a new product?
- How much of the presuppositions is correct, about the role of experience, the role of general human characteristics, human capabilities, limitations?
- ...

Of course these study questions may be deviated from. There are too many to sort through properly in a limited period of time, and this series of questions is by no means exhaustive. In any case do not get hung up about details, like a question for each designed/assumed material or functional instruction for use (see §1.4). Also remember that a use study is much too time-consuming to find out whether a new design, such as a prototype, does indeed function: you can do this yourself, you don't need test subjects for that.

Be aware that something is wrong with an unattached presupposition, in other words, a presupposition that does not return directly in a study question, or indirectly via the study structure. Equally, something is wrong with a presupposition that is not explicitly mentioned, that is smuggled in via a study question or via the study structure, for instance the question whether older people act differently than younger people, or men differently than women. A plausible substantiation must have been given for this presupposition, in other words, why this would be a relevant distinction for the actions and omissions of users. Incidentally, such a plausible substantiation is a mere illusion especially when it concerns age or gender: a difference in age, being of either gender – these are at most so-called *proxies* for characteristics and customs of people that may be directly relevant to how they operate. In general terms, older people have more trouble with physical exertions than young people, not because they are older but because all kinds of physical capabilities will, on average, decline as the years go on. In the same way the gender distinction in different cultures often goes hand-in-hand with having different experiences with the use of products.
In a use study you can try to relate diversity in use to differences in experience, to differences in physical capabilities. As a designer you can benefit from such insights. This is not likely if you did not look beyond age and gender. Such demographical characteristics, also consider profession, can only be useful as *proxies*, for the selection of test subjects, such as when you are looking for relatively strong and weak test subjects, or people who presumably have a lot of experience in

something versus people with little experience in the same thing. Whether a certain characteristic is a *proxy* for a characteristic or capability of people that is design-relevant, is generally based on an assumption. And this assumption need not be correct, certainly not in individual cases.

2.4 Operationalisation: from study questions to study structure

☞ With regard to terms used in study questions they must be possible to research, in other words, it must be possible to operationalise them. This means:

!

a the choice of research conditions, such as in people's homes or simulated in a laboratory environment;

b the choice of test subjects, their 'representativeness';

c documentation of the way to get people working, with a model, a functioning product, in other words how are they given the task/tasks, with what instruction(s);

d documentation of the way issues are raised in the study questions, are observed, registered, measured, namely

- use activities: what people observe ('perception' in figure 1.5), how they reason ('cognition') and what they do and omit ('use actions');
- usability of the model/the product, ('performance', 'side effects', 'effort': mental, physical);
- specified experience ('experience');
- the condition or mood someone is in;
- human qualities and capabilities ('characteristics');
- repetition of the use, sequential effects in the case of more than one task.

In the following paragraphs the points listed under *a* – *d* will be dealt with one by one. For now, just a note on the type of research that use research is, not because this is relevant for the structure of the research, but especially because such a characterisation makes substantive choices more explicit given existing research approaches.

A common research structure is as follows:

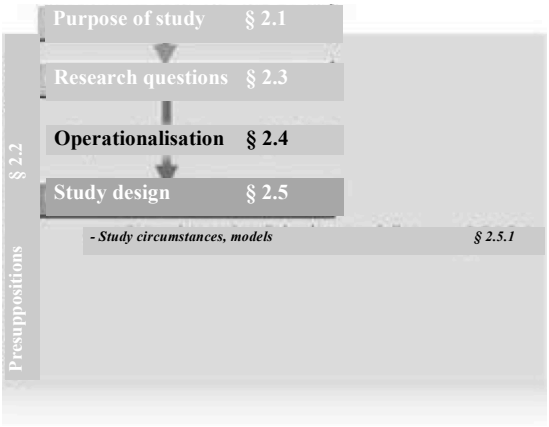
- descriptive research, such as an election poll,
- explorative research, for instance field work in anthropology, and
- hypothesis-evaluating research, characterised by controlled and systematic manipulation of research conditions, consider experiments in biology and psychology.

At the end of §2.1.1 we already commented that textbooks mainly focus on the last category, hypothesis-evaluating research. This type of research is the most 'regulated'.

Use research most resembles the second category, also referred to as hypothesis-generating research, although this classification is less applicable to use research. Use research also has a clear descriptive component: look at the first study question in §2.3, 'What do people do and omit...'. Insofar as use research is evaluating in nature this is most evident in the last study question in §2.3, 'How much of the presuppositions is correct...'. Incidentally, this is a weak form of evaluation because, as a rule, control and the systematic manipulation of research conditions are in fact impossible in use research.

2.5 Research design

2.5.1 Research conditions as natural as possible; models, unobtrusiveness



! Usage studies are performed because you are *lacking the knowledge*, because usage activities of people are simply hard to predict. The best thing for this would be to catch people in their natural environment, in other words, environments which they perceive as being normal without disturbing elements, whilst believing that they are not being watched. Why this natural environment? Because you wish to make statements about use within this context and as you are unaware of the effect that deviations from a natural environment may have on the results. You must realise that extrapolating observations made in a simulated environment into practice just like that indicates a logical error. This is because you then assume that these observations are invariant for differences in environmental elements between the simulated environment and a natural context. However, these clear-cut insights are generally not available at all. If they were, you should really ask yourself whether the amount of knowledge available would not eliminate the need for the entire observation study.

! A similar situation arises on the product side, with regard to the question of whether a model or prototype for an observation study should really include every detail. In fact, this question is already wrong from the start: you are unaware of the details, as this requires the type of insight that is not available and is the exact reason why a usage observation is performed. A simulated study will not tell you anything about which elements to include in a model or prototype. And as long as you are unaware of which aspects of your design are significant to users, you will remain unaware of the distortion that may be the result of omitting design characteristics from a model or prototype.

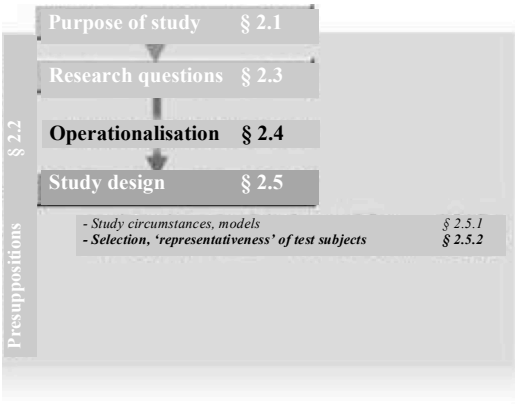
All of this does not mean that a plastic flower on a table will make a difference or not. The important thing is: the more natural/complete the study environment is, the less reason there is for doubting the justification of generalisations, in other words, the smaller the risk of distorted statements when extrapolating results towards practice, in short, the smaller the amount of explanation required. Especially with models and often also with prototypes not much can be said about a natural context. Please note that this limitation will not have the same significance for every product type. Interaction that largely takes place in the heads of the users, for example, information transfer via monitors, appears to be easier to imitate at an early stage compared to interaction whereby physical factors also play an important role, such as the transmission of force.

Plus the usage researcher will also be present as an unnatural element, as researchers in people's homes, who often also have equipment with them, will always be a visible element within the environment, however much they try to keep in the background. One way to tackle this, at least to a certain degree, is to design the study based on unobtrusive observation, by diverting the attention of test subjects, directing it to something else rather than working in an obtrusive study atmosphere. One example is that test subjects in the study into the use of the new condensed milk tubs (Figure 1.3) were offered a cup of coffee even before anything had been said about the study. As a result of this, the use of the tub was observed for half of the test subjects (those who wanted coffee with milk) without them realising it themselves. These test subjects required more attempts to pour milk into the coffee than the rest who did it knowingly in front of the camera (Kanis and Wendel, 1990).

Another example of unobtrusive observation is the usage study into the electric toothbrush.The test subjects did not know that it was about inserting the brush part, nor that it was about the functioning of the bending mechanism. They were left to believe that it was about the brushing of their teeth itself. Apart from this they were brushing in front of a two-way mirror behind which a camera, which they could not see, was pointing straight at their mouths. It was said in advance that recordings would be made, but it was not said how. Naturally the images had to be viewed by the test subjects afterwards for approval. If any of the subjects were to consider the images too intrusive on his or her privacy, the recordings would be erased there and then. None of the test subjects, however, considered this necessary.

More examples will follow below. It is self-evident that an unobtrusive study method will be more difficult at an earlier stage of the design, for example, an initial model that hardly functions at all.

2.5.2 Selection and 'representativeness' of test subjects



Should you always have people from *the target group*? You should if this target group is defined according to characteristics that are directly related to the use of the model or product involved, for example, a product specifically for women or children. However, a target group can also be described in other terms, such as available free time or income. If you have enough test subjects, you may decide to use people with a certain amount of free time or income they can spend: you never know whether, for example, they have a particular personal characteristic, etc., see §2.6.1. However, as soon as the presuppositions give you reason to recruit test subjects with and without a particular characteristic, these target group descriptions that are not related to the use should be abandoned immediately if it turns out to be difficult to get hold of people with the desired variation in that peculiarity.

! These considerations alone show that a usage study does not have to be based on a random sample at all. Quite the contrary, it is generally recommended to work selectively when trying to attract possible participants. The aspects on which selection is based depend on the presuppositions/study questions. Try to achieve as much variation as possible in the characteristics on the basis of which the test subjects are selected: having equally strong test subjects if you think that force is significant will not yield any useful results. Please note that this involves a practical difficulty, as selection on certain characteristics may result in extra emphasis on the study context. One example is experience. If you were to spend hours researching this, in order to get test subjects with one particular experience or other, or no experience at all (assuming this is possible), the end of the story would be that the test subjects would have significant knowledge of the subject of the study at the start of it, plus a number of opportunities would be lost to perform the study unobtrusively.

In a case like this you will have to work with casual indications of large differences in experience ('do you use a computer at all?') and see how this turns out in the study. Remember that usage studies are not experiments with accurately described study conditions, even though experience in particular tends to vary widely from one person to the other: selecting people with 'the same' level of experience is generally impossible.

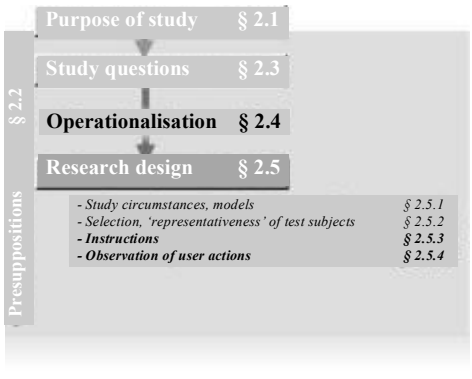
Observation of children as test subject requires a number of special precautions. When applying the Dutch standard for the testing of childproof packaging (NEN, 1983) children are given five

minutes to open a packaging. The exact instruction method is not described in the standard. The following guidelines are, however, given:

- place the children in a trusted environment;
- also ensure that distractions are minimised, literally: ‘keep them away from regular toddler/nursery activities’;
- limit the size of groups of children to three, ‘compile the groups in a manner that will be familiar to them’;
- no parents.

☞ And then there also is the term representativeness. The usefulness of this widely spread jargon is minimal if people fail to mention: representative for what, on what characteristics. Without adding this the decorative term representative means nothing. But even if it has been specified – for example, according to age, gender, body height, annual income – it may still be a meaningless qualification. For example, what is the use of knowing whether the anthropometric characteristics of a certain group of test subjects can be used to estimate these characteristics for the entire population from which this group was taken with a particular confidence interval if these anthropometric characteristics appear to be in no way related to the subject of the study, for example, whether or not usage issues will arise? The answer is: nothing at all, in this respect. Characteristics that, according to the analysis, are related to issues like these, for example, whether someone has a particular level of experience or not, do not have to correlate with the anthropometric characteristics mentioned above at all. In short, useful statements about the representativeness of test subjects can only be based on knowledge of the (distribution of the) characteristics that are related to the phenomenon studied. However, the question of whether these relationships exist is often still a matter of study and they are nothing more than assumptions when test subjects are selected in a case like this. Sometimes you have more certainty, and it is an obvious blunder to select (only) industrial design engineers as test subjects, for example, for a usage study into a new software package for home banking that is intended for use by anyone.

2.5.3 *Instructions and unobtrusiveness*



☞ It speaks for itself that the instructions should be aimed at the test design and should therefore yield results that can be used to answer the study questions. Remember that the instructions should not inadvertently include matters that are not discussed at all, for example, the effects of a product on different materials, unless these new matters are used to provoke test subjects, to distract them from the actual task (unobtrusive observation).

!

Uniformity in terms of content is required for the instructions, to give every test subject the same starting point as much as possible. The instructions should therefore be written down from the start. The more explicit instructions are, the more the emphasis will lie on the study context. If it is not required, never mention every step separately, in terms of ‘could you start unpacking now’, ‘could you make it ready for use’ (consider e.g. the inclusion of the words ‘ready for use’ in instructions), ‘could you now please use it to...’, etc. Where possible state an overall task in one line that focuses on the results of the use instead of this; use normal colloquial language without hoo-ha and without jargon.

If you want a product to be picked up by test subjects first, place it in a corner to force the test subjects to do this without you having to tell them. If you want them to do something like replace a blade, give them the product without a blade or with a blunt blade, keep the new blades somewhere close, in an obvious place, so that you will not have to tell them everything and will be able to observe unobtrusively. Where possible, provide products that are in an awkward, unpractical state so that something you want to learn more about must be done with them automatically – setting too low, empty, wrong default setting. Consider whether test subjects can be put on the wrong track in a useful manner, in a similar way as for the toothbrush (see §2.5.1). For models that are not fully functional, explain briefly and clearly which functionality has not yet been implemented in the design. In other words, no riddles and no going into all those great design intentions that are still on their way but have not been realised yet. And again: the more natural the tasks, the smaller the amount of explanation required afterwards when extrapolating to practice. And, as a rule, the study should be designed in such a way that required instructions are minimised.

This leaves the manual. You cannot get away with characterising a manual as a shortcoming of a designer, as if every product should be able to explain itself entirely in terms of functionality and use. Think about the almost endless possibilities of electronics, which designers can only partially implement in their designs. Just like the manual of many fax machines has quickly run into dozens of pages, the manual of a mobile telephone will not be that much shorter either, unless the instructions for use are integrated in the software, for example, in so-called active help functions (‘embedded’); see, for example, Davis 1998. As long as the performance of usage studies does not involve the testing of, for example, the clarity of a manual, it is often better not to make it available to test subjects: it contains a lot more than what is covered by the study, which makes it more difficult to find the required information, and that information, in so far as it is correct, may simply be too detailed and may give away the subject of the usage study. In a case like this it is important to prepare a brief manual yourself for the test subjects with the information that can be deemed essential and just sufficient.

2.5.4 Observation of usage actions

Video recordings are an obvious means (see for a detailed discussion of this Jordan and Henderson, 1995); a camera is often still a better alternative to just pen and paper. To what

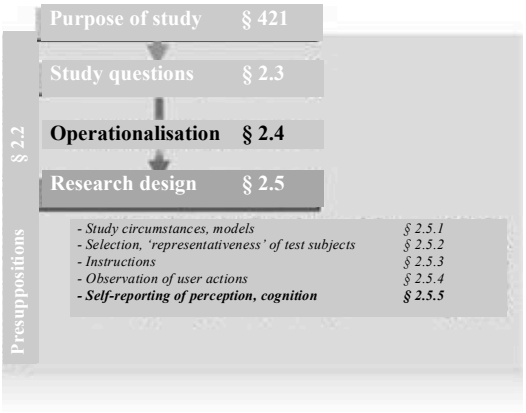
extent does the obtrusiveness of video equipment have an adverse effect? Several studies, including studies covering just the effects of sound recording equipment, show that test subjects often have the pleasant tendency to forget about the presence of recording equipment after a while (Weick, 1968). In any case, it is a huge misconception to use the limited presence of pen and paper as an excuse to promote the use of crossing-out lists, or rather, checklists, sometimes even as an alternative to video. It is naturally sensible to consider possible results before carrying out a usage study. But it shows remarkable stupidity if someone believes that collecting data in a usage study is simply a matter of ticking categories that have been thought up behind a desk. If that were possible, you would evidently have enough knowledge already to eliminate the need for a usage study completely: after all, it is about finding out which usage methods exist, not their frequency. An additional disadvantage of imaginary certainties about what test subject will (and can) do is that expectations like these may steer the observations of researchers, causing them to miss things that do not match their ideas. As the poet said: he who finds something has not searched enough. In other words: you find what you expect and stop looking; refer to the self-caused short-sightedness as a result of 'over-usecued/over-presumed' usage studies (see §2.2 near the end). Then there is the 'higher' form of study: seeing what is not done, listening to what is not said. This is without mentioning the impossibility to keep track of any decent checklist in practice: far too much is often going on at the same time to manage that.

One thing that may be very useful instead of a checklist is for one of the researchers to join in the observation, for example, via a monitor, and write down certain notable events and actions, so that they will not forget to discuss them with the test subject afterwards to find out the exact difficulty, etc. This does, however, require a minimum of two people.
The following concludes this paragraph. The recording of user actions on video can sometimes be overdoing it. This was the case for a study into creating a special voice mail for a telephone. It appeared that the procedure followed by test subjects, in other words, what they entered and in which order, could be tracked very well on the basis of the flow chart depicted in Figure 3.7. Another reason to keep the video running anyway is sound registration, for example, if people think aloud, see next paragraph. For the creation of the voice mail via the telephone nobody was thinking aloud, as this would have interfered with adequate reactions to instructions issued via the telephone for the creation of that voice mail.
It is generally recommended to avoid collecting data that may be considered superfluous based on its nature and/or details on the basis of convincing arguments. Quite often this does not only
☞ save time and money, but it also prevents you from losing track during the analysis.

2.5.5 Self-report about perception and cognition: 'thinking aloud' versus explanation afterwards

Actions performed by users can be seen directly, but not what users observe and think, whilst that information is exactly what may help designers. You can make users say aloud what they

think. If you make users do that during use, it is also called thinking aloud. You can also ask questions afterwards.



In usage studies thinking aloud usually yields weak protocols (= literal reproduction of what a test subject has said), see Figure 2.2 with part of a protocol for the use of a sphygmometer (Rooden, 1998). Parts of the interaction with common products are so familiar to the users that they are performed automatically. They will hardly say anything about this. For the rest of the interaction perception and cognition can be put into words if the test subjects remember to do so, deem it useful and can come up with the right words. Test subjects may 'forget' to talk if they are too busy with the task or because they are used to working in silence. Ironically, users fall silent when they run into trouble. That is inconvenient, as it means that information about possible causes for the problems are missed. It may also be that a user thinks too much to put everything into words. The user will then make an intentional or unintentional choice and say what he or she deems relevant to the study. There may also be things that test subjects would rather not say, for example, to cover up 'stupidities'. It may help to explain the purpose of the study clearly beforehand and create an atmosphere of trust. Test subjects have widely differing verbal capacities. Some matters, such as manipulations, are difficult for many people to put into words. Thankfully actions can be observed directly. If you show a drawing of a new design to users, it may be a problem that limited vocabulary will result in a description that is too general. You can then encourage users to point at things and make gestures in the air. To hear more from test subjects you can also ask questions afterwards or play a more active role as test leader during use. It may be preferable to ask questions afterwards, as you want the use of the product to be as natural as possible. This may be important, for example, if you want to know how much time people require. If you ask questions about perception and cognition, you will not be sure whether it was actually experienced during use. This is because the test subject may reconsider things after having used the product. Perception and cognition experienced during use may disappear from memory or be distorted. Test subjects may try to justify their behaviour and they may start to show expert behaviour. They will then start to speculate which problems other users may have.

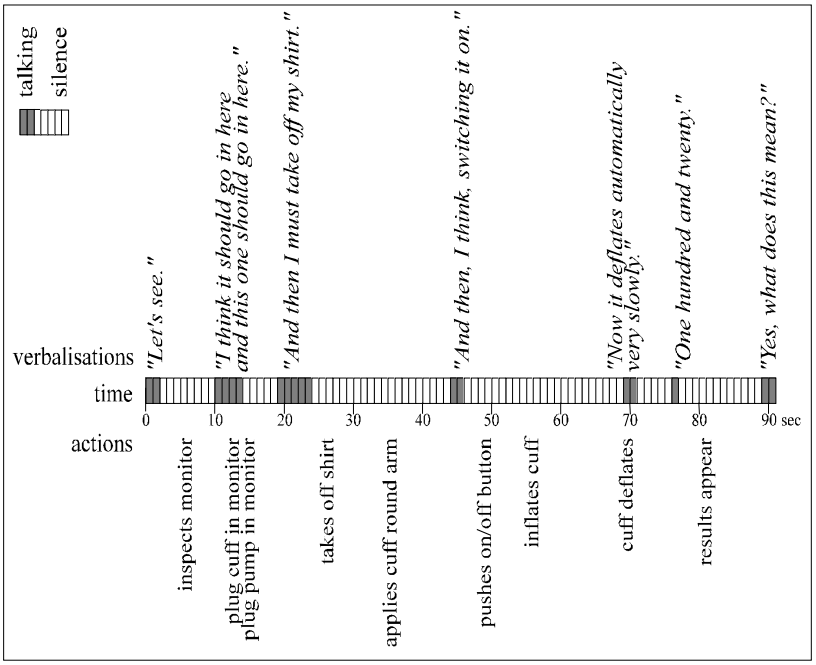


Figure 2.2 *Ninety seconds of a protocol for the use of a sphygmometer*
Nothing is said for most of the time. And whatever is said mainly concerns the actions that can already be seen on the video tape. Hardly anything is reported about perception and cognition.

Both examples below are taken from the study by Rooden referred to above into the use of a sphygmometer. The questions were asked whilst viewing a video recording of the use together with the test subject.

[question by researcher] 'I can see that you put the cuff around your left arm, did you have a reason for that?'

[answer by test subject] 'No. When I visit a doctor, they usually say: please extend your left arm. But it doesn't really matter. It is easy though, if you are right-handed, as you will operate the pump with your right hand. So you put the cuff around your left arm.'

Comment: This sounds like trying to make sense of things in retrospect. The answer starts with a no and is then followed by reasons to support the choice for the left arm. If this line of reasoning had actually taken place when the user chose which arm to use, you would expect some hesitation.

[question by researcher] 'Why do you think that you didn't remove the air from the cuff first?'

[answer by test subject] 'I don't know, it was not very wide so I didn't really notice. I mean, if it had slipped down I would have noticed earlier. But it was already quite tight, that's why.'

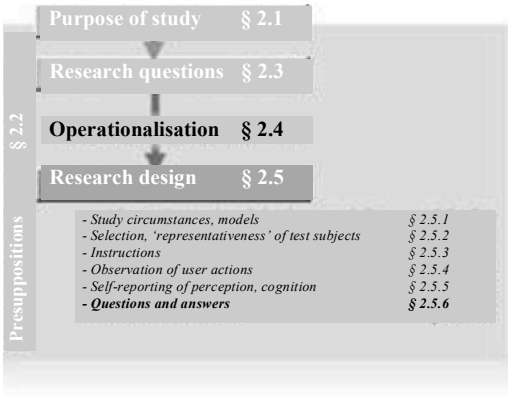
Comment: If an answer starts with ‘I don’t know’ you can bet that the rest of the answer has been thought up afterwards.

Negative statements, for example, about matters that have not been observed or understood are not subject to the suspicion of distortion and are therefore often more useful. It also sounds very good if users sometimes say that they can’t remember what they thought during use. However, information gathered afterwards may be useful in a design process, especially if the information gathered by thinking aloud is limited. See also the next paragraph (§2.5.6, *Questions & Answers*).

The risk of distortion of information gathered afterwards is minimised when the event is remembered well by the user. The best thing to do is ask your questions immediately after use or refresh the memory of the user by showing a video recording of the event and base your questions on that.

! If you ask users questions during use or encourage them to keep thinking aloud, you have a greater chance of extracting information that is actually relevant during use. However, an active test leader like that can affect the use. As long as your aim is to observe plausible usage methods, this would not appear to be a problem in a design process. There are, for example, other ways to keep test subjects talking, which are less obvious than reminding a test subject of it all the time. In ‘co-discovery’ two test subjects discover a new product together. The idea is for a natural dialogue to start between these two people, whereby they tell each other what they think should be done; please refer to the studies by Suchman and Frolich et al. mentioned in §1.7.

2.5.6 Questions & answers



The aim of encouraging test subjects to make statements is, for example, finding out how people perceive the use shown and why. In usage studies, you should try to make test subjects report matters spontaneously: this will then concern usage aspects that are apparently important to them. As a researcher you should exercise restraint here, in other words: have the test subjects state matters themselves as much as possible without allowing them to be influenced by you. You should therefore never interrupt a test subject and immediately stop talking when you are interrupted yourself.

What people do does not have to match what they say about it at all when asked: there is always a risk of lip service and smooth talking, see the example given above. In this respect the convenience of a checklist clashes completely with the usefulness of usage observations. This obviously does not mean that questions should be avoided in a usage study. It should, however, be noted that a lot can go wrong with them. For starters you should avoid biased questions: ‘Did

you find it convenient, comfortable, difficult...? Does the product have a convenient handle? Do you think it can be used easily? Are you able to work precisely enough with the prototype?' The bias lies in all the adverbs and adjectives: try not to use these.

A good alternative is presenting a subject with a question like 'How was...' or 'What is your opinion about...', followed by questions for clarification, that is to say, on the basis of which an experience is called positive or negative by a test subject. In other words, the trick in a case like this is to find out the internal references of the test subject. What doesn't help is asking test subjects to assign some kind of score, for example, using a scale of 1 to ... . This will generally only yield what appears to be 'hard' data but isn't, for example, if scale references in words are interpreted differently.

We will now discuss a number of findings relating to 'questions and answers' taken from literature. In two studies (Bercini, 1992; Foddy, 1998) the following reasons were given for the doubtfulness of answers to questions posed to test subjects:

a people interpret key concepts differently, they may not understand them;
b alternative answers provided insufficient (only for closed questions, which are generally not used for usage studies);
c the frame of reference for answering questions differs between people, who may have any number of different starting points, also called 'perspectives';
d people may have trouble remembering relevant aspects, leading to incompleteness, hit-or-miss answers;
e effect of social acceptability, suspected risks of answering.
Categories may overlap, especially *a* and *c*, and *e* with all the others.

So what are indications that something is (or may be) wrong (see especially Foddy)? If people cannot repeat a question word by word or only partially. And if people indicate that they are not sure about their answer when asked about it. Fast, plausible answers are definitely no guarantee that it is fine to ask a particular question, as asking the same question to different people basically means that the same question is answered, with the same level of completeness, without conjecture, etc.

Examples:

(re *a*) (what do key concepts mean, how are they interpreted?)
'In the past year have you been bothered by pain in your abdomen?' (Bercini)
Most of the respondents (twelve) could answer this question straight away. However, each of the twelve respondents indicated a different part of the body as the abdomen when they were asked. Additionally, further questioning revealed that, in terms of 'to bother', some would not be that worried about the pain unless they were aware of its cause. For others, 'to bother' has to do with the severity of the pain: they would only have answered 'yes' if they had had severe pain.

(re *a*) (what do key concepts mean, how are they interpreted?) and ad *c* (frame of reference, starting points)
'How often have you had to stay in bed for at least half a day due to illness since 1 January?' (Foddy)
How long is half a day: = 3 hours: 11%; = 4 hours: 46%; = 5 hours: 16%; = 6 hours: 18%. For this question 14% of the respondents was thinking of actually being ill and 10% was thinking of staying home from work. Foddy concludes that people who are asked the same question in words may actually answer different questions. He also refers to the urge of people to answer (p.103).

(re a) How long have you used the device? (Bercini)
This question was about a number of aids, such as wheelchairs, walking sticks and hearing aids. When asked it appeared that many respondents used their aids every now and again, sometimes they didn't use them for a long time. Some answered the question with the actual (estimated) time they had used it, others stated the number of years since they had used it first, whilst yet others thought that the question only referred to how long they had owned the aid that they were using at the time the question was asked (and therefore not how long they had been using this kind of aid).

(re *b*) (response options insufficient)
'Which of the following types of food increases or reduces the risk of cancer when eaten regularly? (incl.) ... wholemeal bread, semi-skimmed milk, whole milk and cheese, popcorn, bacon and eggs, fresh fruit, fried food, fruit juice' (Bercini).
Respondents turned out to have little trouble making the correct link between food and cancer. But when questioned about it in more detail, it appeared that this was not really based on actual knowledge (about fibres, fat content), but more on guesswork based on general notions about healthy food and unhealthy food. In other words: you will be very mistaken if the answers to a question like this lead to the conclusion that people have considerable knowledge about carcinogenic food.

The following text is taken from an article by Steenbekkers (1998) about a study into anthropometrics, capacities and activities of elderly people.

> 'When further questions were asked, it also appeared that some people who had indicated not to have any trouble performing a particular activity on the 4-point scale never performed this activity in daily life. Men, for example, regularly answered the question relating to having trouble hanging up large pieces of laundry by saying that they never had any trouble doing this. The verbal explanation, however, revealed that a lot of men never perform this activity, as their wives always did the washing. They usually then added that they expected not to have any trouble doing it. They also regularly indicated that getting in and out of a bath was trouble-free, even though further questioning showed that they only had showers. Further questioning also showed that people use a wide variety of compensations in daily life for performing certain activities despite physical complaints or reduced capacity. The interview regularly states examples concerning, for example, the opening of a jam jar. Examples of tricks they use are: tapping the lid or the bottom with a knife; holding the lid in warm water or on a warm hob; piercing the lid. In addition to this, a large number of tools are used, from pliers that are intended for this or pieces of cloth to sandpaper or rubber household gloves. Others are also asked to open a pot that has been vacuum-sealed.'
> The fact that all sorts of compensations and tricks are not revealed until further questions are asked shows that the interpretation of questions by respondents is not always what the person asking (or preparing) the questions had in mind. Language simply cannot be used to record things unambiguously [refer to what is said about this in §1.4.2 and §1.6.1].

(re *c*) (frame of reference, starting points)
'How would you dispose of the following items if you had to: paint (dissolver/thinner), pesticides, herbicides, waste engine oil, batteries?' (Bercini)
For starters, this 'environmental question' turned out to be irrelevant for items that people never want to get rid of. Apart from this, there was a difference between what people said

they should do and what they thought they would really do, depending on the circumstances, for example, the amount of paint they had to get rid of. Answers with no strings attached to questions like these are easily coloured by social acceptability.

(re *d*) (what can people remember?)
People were asked if they knew whether any adult family members had had a tetanus shot in the past 10 years and, if so, how long ago this had been and whether it had only been a tetanus shot or a tetanus-diphtheria combination (Bercini).
Respondents were often unsure about tetanus shots given to other family members, which initially resulted in the answer 'no' and therefore in poor reporting. Respondents were also unaware of any combination shots. In short, this was asking far too much in terms of what people can reasonably remember.

(re *e*) (social desirability, rationalisation)
'Why haven't you tested your house for radon?' (Bercini)
This question is based on the assumption that the threat posed by radon is generally recognised to a certain degree and that people therefore have logical explanations for why their house does not have to be tested or has not been tested yet. People responded to the question with socially acceptable, rationally sounding answers. However, further questioning revealed that many respondents had never thought about it for one second before the interview and therefore came up with instant rationalisations.

See also the following example taken from the study by Rooden (1997) into a sphygmometer.

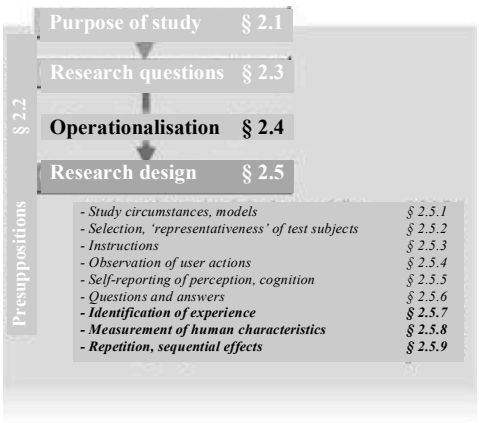
[question by researcher]'Are you normally speaking left-handed or right-handed?'

[answer by test subject]	'I am right-handed and all my injections have been in my right arm. And I row on the starboard side, I am very consistent.'
Comments:	Even though this participant indicates that he is right-handed, he still pumps with his left hand during the test. He piles up a number of things that are related to being right-handed in the answer to this question. Is he trying to come across as convincing?

Anyone who wishes to learn more about the significance and expressiveness of answers within the context of questionnaires can refer to studies that are known as 'cognitive interviewing'. In these studies questions for surveys are analysed to see if they could have possible tendencies (see for example Campanelli, 1997; Willis and Schechter, 1997).

2.5.7 Identification of experience

When asking an experience-related question, don't stop when they say that they have it or not. Ask which product is involved exactly, which component and why they think of that of all things/why this is considered to be similar. Remember that the relevance of experience stems from applying situated knowledge, in other words, knowledge that can be used in those circumstances. Present the model, prototype, product involved, to allow test subjects to point out components, to make statements as tangible as possible.



The ideal situation is if test subjects can also demonstrate experience with a product of their own which they mention, if they have it with them. In usage studies at people's homes this is generally not a problem. For observations in a room specifically designed for this (for example, the usage laboratory at IO) test subjects would be required to bring along (parts of) products which they consider similar. This is often impossible for practical reasons. What's more, you don't know what people consider similar, whilst asking all kinds of questions about this in advance is wrong due to the possible steering effects of them, please refer to what is stated about this in §2.5.2, about selection and 'representativeness' of test subjects. The *type* of experience is generally much more important than the intensity, how often something is used, see also the analysis (§2.8).

2.5.8 Measurement of human characteristics

Measurement of maximum forces to be exerted. Hand dynamometers are generally available for this. These can be adjusted to hand size. Instruct the test subject to increase the force gradually, in other words, not with shocks or explosively, and to hold the maximum force for a number of seconds.

Measuring hand size. This can be done, for example, by making a recording with the video from straight above of the hand flat on a mat with a ruler (these mats are generally available).

Left and right-handedness. This characteristic may differ for various activities, such as writing, throwing a ball, lighting the gas. As a result of this, people are left or right-handed for certain tasks and not in general.

! Always take measurements, for example of forces, at the end, after the usage observations, which are what the study is done for in the first place, again to prevent unwanted steering by experiments; refer to what is stated in the previous paragraph about experience-related questions.

2.5.9 Repetition, sequential and/or pass-on effects

! The ideal situation is to repeat tasks at least once, to be able to say at least something afterwards about:

- learning when people are confronted with something new, in other words, when observing initial use (this is quite often a weak side of usage studies, refer also to one of the study questions in §2.3) and
- variability within subjects compared to that between people, in so far as learning does not appear to be involved: as a rule the variability between subjects is considerably greater than

that within subjects, however, the latter may still be significant, see Weegels (1998) about the use of household cleaning products.

Try and spread out different performances of the same task as much as possible over time, have a different task performed in between, where possible, instead of the same task again and again. This does, however, give rise to a new problem, namely sequential or pass-on effects for successive tasks. This may involve the same task performed with different products, but it may also involve different tasks performed with the same product. These effects only occur in a so-called 'within subjects' design, in other words, when the test subjects are asked to perform more than one task. A great advantage of this study design is that you obtain observations which can be compared intra-individually, i.e. for each test subject. In this design you will require far fewer test subjects than when you have each test subject perform only one assignment – this is called a 'between subjects' design. The latter study design is not subject to sequential effects, as every test subject is given his or her task as the first and only one.

☞

☞ Pass-on effects occur when people gradually learn during a study, they are trained, they may get tired, they may lose interest. Basically it does make a difference whether each task is assigned as the first, second, etc., equally often. This is because an essential aspect is which task/tasks preceded it. However, even if you make sure that the frequency of the different histories – for example, performing task a and then task b as often as the other way round – is the same, things may still go wrong. This is because the effects of how users proceed when performing task b after task a may still affect what they do when they perform task b first and then task a.

You often read in study reports that matters were 'randomised': everything is mixed up as it were, after which it should all average out. This does not have to be true. See the analysis (§2.8) for how to tackle this further.

2.6 Criteria for findings during studies (validity, etc.)

2.7 Performance, pilot

2.7.1 Overview of points for attention relating to use

Make sure that you enter the study with an overview of points for attention relating to use which must be as complete as possible, in view of the presuppositions made and the study questions formulated. You should gather information from all test subjects for all of the items in this overview before they leave.

Depending on how the study is designed, this overview of points for attention relating to use may consist of:

- all designed/presumed material and functional instructions, to determine whether they have mediated, helped, put people on the wrong track..., see table 2.1 (again: usecues may be entirely superfluous, someone may do something correctly automatically; what's more, it always concerns usecues within context, as opposed to isolated, individually operating cues);
- elaborating on problems as a result of difficulties observed, for example, whether the user can work accurately or whether enough feedback is provided, all of this then related to (functional) product characteristics (a general remark such as 'no, it didn't go very well, I had trouble performing the task' is useless to a designer);
- similar if people have trouble with a particular task, see Figure 2.5; perceptually (to notice it), cognitively (to understand the meaning of it) and physically (to actually do it);
- the reason why certain usage tasks are performed (in a particular order);
- experience with other products and why these or components of them are considered to be similar and in what way;
- .....;
- without fussing about alleged buying behaviour, about beauty or ugliness.

Make sure that test subjects speak for themselves first, about their own experiences and do not lapse into 'expert' behaviour, in other words, state what mainly will happen to others if, for example, they do not understand things (no, a test subject will is of course fully aware – what were you thinking?).

After the usage observations, discuss usage aspects from the overview that have not been covered yet; this is to check whether the test subjects really have nothing to say about them. Try and understand what test subjects mean in connection with what you have seen and heard them

do, also remember the notes made by fellow researchers relating to special events, see §2.5.4. Never work with ready-made questionnaires. Always be fully aware that you may regularly run into something that hasn't been taken into account at all during this type of observation study. An unexpected observation like this may provide an insight that you may want to share with the other test subjects. In other words: if this had been thought up in advance, this insight would have been included in the overview of points for attention relating to use from the start. This overview can therefore never claim to be exhaustive, the actions and omissions of every successive test subject may be a reason to add one or more points for attention.
So how to do this then, especially in view of comparability (see §2.5.3, instructions)? Let this be clear: do not bother test subjects again that have already been there. This means that you will only have the next test subjects. It should also be clear that it would be madness not to confront the next test subjects with something that you are very keen to know, that you decide not to do this out of inflexibility, only to maintain full comparability between all the test subjects. Tying your hands like this – not changing anything once the study has started – may be necessary for experimental studies, but this would result in completely undesirable limitations for usage observations. However, in usage studies the comparability between test subjects is often important as well. To maintain this comparability as much as possible, a new point for attention which is identified during the observations should be presented to the remaining test subjects at

! the end of a session as much as possible. This means that the comparability between the test subjects is retained for all experiences preceding it.

In short, the overview of points for attention relating to use may grow during the study, it may become more specific. Finally, watch out for the following. Something is wrong if the overview of points for attention relating to use contains other/new questions compared to the study questions. This means that you cannot consider those other/new questions in the overview of points for attention relating to use to be an elaboration on or specification of the study questions asked before. In this case something must be done: either the study questions must be changed, or those other/new questions in the overview of points for attention must be removed as superfluous.

2.7.2 Who should perform the study?

The designer is both the most vulnerable researcher of his own idea and the most obvious one. The most vulnerable one, as it is often hard to watch all the horrible things those users do to that beautiful design. Quite often it turns out to be highly difficult for designers to maintain a distance, to stop themselves interfering when a user does something stupid again. On the other hand, it is most effective if a designer involved is right there on top of things, that he or she does not have to hear things from others, see also Freudenthal, 1999, p.229.

You will learn most by being the devil's advocate. If in the report of a usage study with design models that were performed by the designer(s) involved everything appears to fit the

presuppositions perfectly, this is a reason to be highly suspicious. In usage studies into an existing product of unknown origin that must be redesigned, there may be less reason for suspicion if only pleasant aspects are revealed.
In that case never present yourself to test subjects as the designer; remember the disturbing effect of social acceptability that makes people say what they think the researcher wants to hear, makes them say things they think will make them look good. It's better to say that the design was made by another group or by the competition.
Whatever your task within a usage study, the rule is that you will gradually become more skilled in it. There are many reasons in favour of not changing a division of tasks agreed between researchers at the start, of maintaining it throughout the study. But it also concerns studies involving a learning process. This learning to perform a study may, on the other hand, be a good argument to vary study tasks between the people involved.

2.7.3 Intervention

Do not help too soon if a test subject gets stuck. Let them muddle on for a while before asking questions like 'Well? Can you manage?', followed 'What's the problem then?'. Only give a hint if the problem is really obvious and there is no chance of the test subject finding it out on his or her own, but limit it as much as possible, for example, 'Have you seen that?' or 'Think about this...'.

2.7.4 The end of a session

Imagine the recording equipment being switched off, as the test with a test subject has finished. They talk things over for a while, whereby the conversation quickly returns to the study. While the researcher is already holding the doorknob, the test subject presents a number of new aspects that are highly essential to the study, but may be lost if they are not recorded straight away. This is why recording equipment should preferably be left on until a test subject has actually left.

2.7.5 Pilot

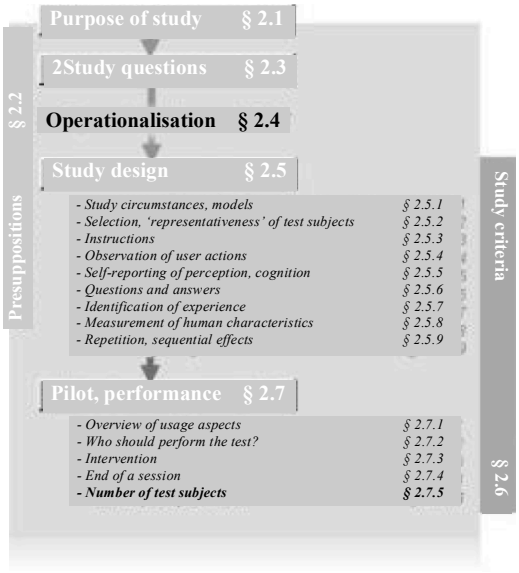
☞ All matters discussed above are relevant or may be relevant in the pilot. It is generally not recommended to perform a pilot with a 'clever boy', someone who knows his way around when it comes to matters of use. It is not only a matter of whether the cameras are in the correct position or the sound reproduction is clear, or the conversation afterwards runs smoothly without biased questions, but also whether the study design is sturdy enough. For example, one design group thought of placing an object with protrusions that had to be cut off with a newly designed pneumatic saw in the centre of a table, which meant that some of these protrusions that had been attached to the object were impossible to remove without turning around the saw blade. The intention is clear: unintentionally (unobtrusively) encouraging test subjects to discover the possibility of turning around the saw blade and putting it into practice. In the pilot this went fine, that is to say, with a 'clever boy'. However, the real test subjects who came next were less in control: the object on which the saw had to be used started to move along with it. There was no other option but to clamp this object to the side of the table. And the result was: the extra

manoeuvring space at the edge of the table allowed the real test subjects to saw off all the protrusions without changing the saw blade, among other things, by holding the saw upside down.

! Immediately after the (first) pilot test subject, watch the entire video to see whether everything was recorded. If anything is wrong it must be corrected immediately, to see if the modification is an improvement for a possible second pilot test subject. As a rule a pilot leads to a number of changes in the study design. Only if a pilot does not result in any changes at all, the results of the pilot may be included in the analysis. The decision of whether a pilot has gone well and therefore the study with the real test subjects can be started is up to you as researchers.

2.7.6 Number of test subjects

In most of the examples given so far (see both previous chapters) the number of test subjects was limited, in most cases even to such an extent that statistical reliability was still a long way off. Does a small number of test subjects now also imply that the significance of what has or has not been observed is limited? Please look at figures 2.3 and 2.4 for this. Figure 2.3 shows the possibility of a particular event, p, as a function of the possibility K_o that you have not observed that event after n test subjects (random sample). As can be seen in this figure, the possibility p must be small or the number of test subjects must be relatively large if you wish to say anything with any certainty about the probability of events that have not been observed. Figure 2.4 shows the other side: the possibility K_+ that you have seen a particular event with possibility p at least once after n test subjects.



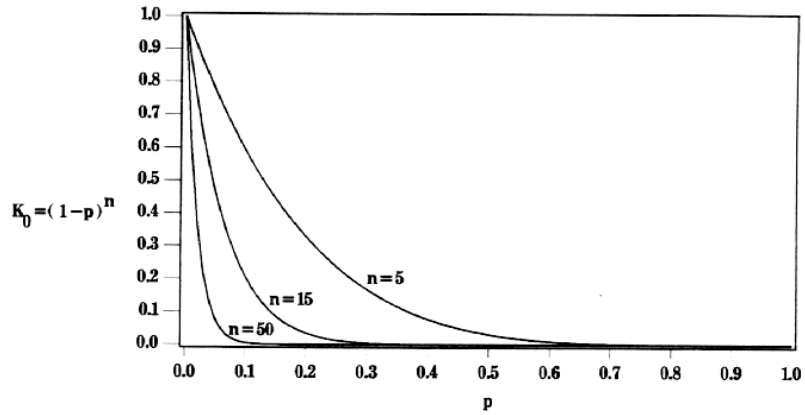


Figure 2.3 *$K_o=(1-p)^n$: possibility that something has not been observed after n test subjects*

!

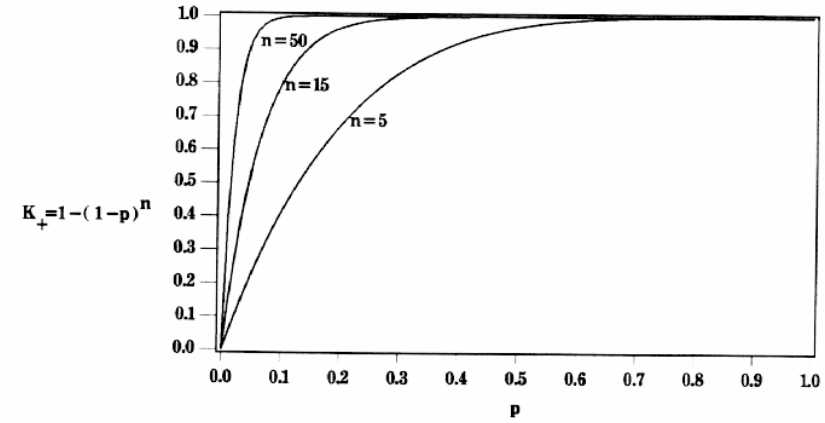


Figure 2.4 *$K_+=1-(1-p)^n$: possibility that something has been observed at least once after n test subjects*

!

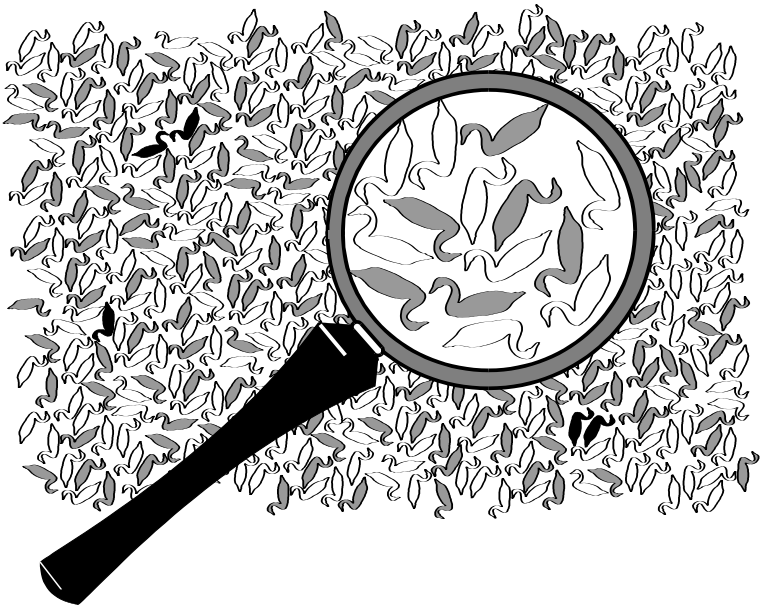


Figure 2.5 *The limited possibility that events found in a small random sample are rare.*

Conclusion: whatever you observe with a small group of test subjects will rarely be highly unusual. In other words: ironically, the limited number of test subjects gives weight to what has been observed. And that is quite different to the well-known platitude that the number of test subjects was once again much too small. See for an illustration of this explanation Figure 2.5.

Let's take the use of the vegetable slicer in Figure 2.6 as a practical example.

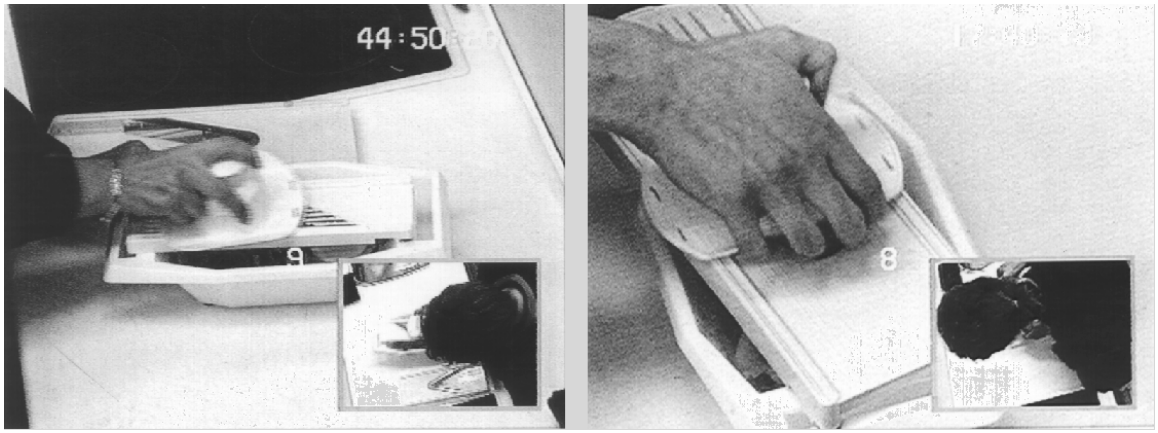


Figure 2.6 *The use of a vegetable slicer. At the left the participant (one of nine) holds a the protective cap from the vegetable slicer steady like the designer (s) this possibly had foreseen. This cap is there for a reason: with this type of vegetable slicer accidents happen frequently, which may have to do with the tapered blades. As a user it is hardly visible because with a common way of working, as shown, your hand is in front of the cap. What the person at the right is doing is thus very dangerous.*

! *One of the total of nine test subjects 'almost kills himself' (to the right of Figure 3.2). The possibility of something like that being revealed in a usage study with* n=9 *(random sample) is listed in Table 2.3. According to this table there is a reasonable chance that between one in five and one in twenty users go about things in the same way as depicted in Figure 3.2, which is rather high in view of the risk of accidents of this usage method.*

Given: frequency f of a certain usage method per 100 users	*Possibility of finding a certain usage method once in a random sample of nine: $9.f/100(1-f/100)^8$*
1	0.08
2	0.15
5	0.30
10	0.39
20	0.30
30	0.16
50	0.02

Table 2.3 Possibility of a certain observation for a given random sample size

☞ Table 2.3 and also the figures 2.3 and 2.4 provide an insight into the link between random sample size and the possibility of observing a particular phenomenon or not, whereby the average possibility that this phenomenon occurs with a test subject is known. In practice this average possibility is unknown. In addition to this, you are generally interested in more than one phenomenon, just think about the various problems that people have, something into which you

!

gain an insight during the study. What you would like to know during a study is the probability of finding something new that is important with the next test subject. This is because the purpose of usage studies is not to establish the frequency of a particular problem (see §2.5.4), but to gain an insight into the various types of problems, as it is much more important for a designer to know that a problem occurs, for example, among only ten test subjects than that one test subject has this problem, or more test subjects. And wouldn't it be nice to have a criterion to avoid terminating a usage study too soon (important aspects missed) or allowing it to continue for too long (waste of time and money).

As soon as it no longer concerns a single problem but more, the formula applied in Figure 2.4 can be written as follows $F_n = F_\infty[1-(1-\bar{p})^n]$, with F_n the number of different phenomena found after *n* test subjects that are important ('*F*' for 'found'), and $\bar{p}$ the average probability of finding any important phenomenon.

This formula is used in literature based on the number of different phenomena found after a large number of test subjects (mostly usage problems) as F_∞, to calculate F_n (Virzi, 1992; Lewis, 1993). This reverse calculation has led to the rule of thumb that four to five test subjects will be sufficient to identify 80% of the phenomena you would like to know. See the excited report about this by Jacob Nielsen (the famous one, the heuristics scientist) on the web (2000). However, in actual fact this rule of thumb is partially based on guesswork. For example, F_n appears to be overestimated, which can be traced back, among other things, to the averaging of the various probabilities for each of the important phenomena to occur into an average $\bar{p}$. You can also put it differently: given a particular F_n, F_∞ is underestimated. In slightly different words than above, the reason for this is that the indicated averaging into $\bar{p}$, underestimates the number of phenomena (such as usage problems) that test subjects have in common, and even more so as the probabilities per phenomenon lie further apart, in other words, if rare cases occur in addition to frequent ones.

!

At this moment the merits are being worked out of an algorithm that offers the possibility to keep track of things with regard to the consideration of whether to stop or continue during a usage study. We will state this algorithm here without further explaining the mathematical considerations – these are complicated and fall outside of the scope of this manual (for enthusiasts: Kanis & Arisz, 2000). The algorithm, which so far appears to predict quite reasonably well, is as follows: $F_\infty = \bar{F}_1 . \bar{F}_{n-1} / (\bar{F}_1 + \bar{F}_{n-1} - F_n)$, with $\bar{F}_1$ and $\bar{F}_{n-1}$ the averages of the number of different cases found for all combinations after *n* test subjects.
Both $\bar{F}_1$ and $\bar{F}_{n-1}$ are averages of n figures: $\binom{n}{j} = n$ for $j = 1$ and for $j = n-1$.

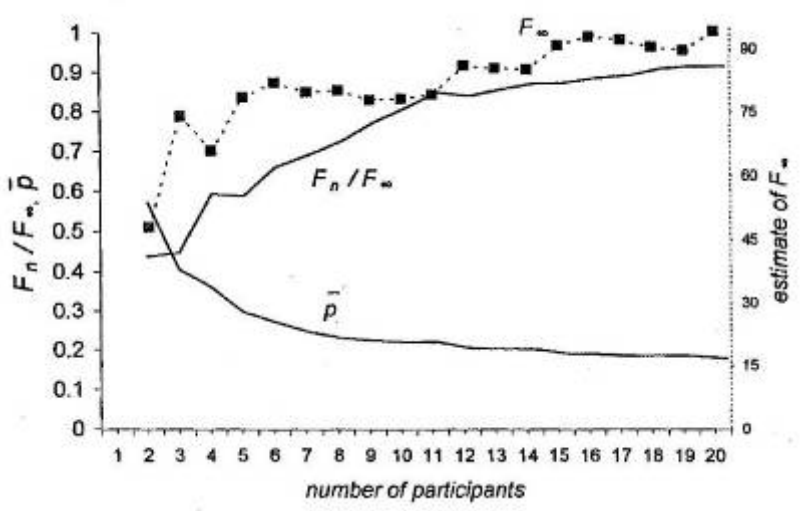
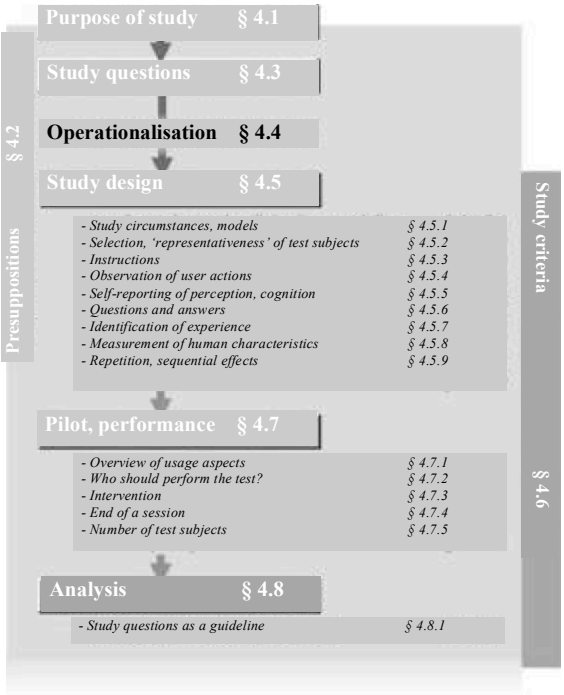


Figure 2.7 *Prediction by experts of problems using a coffee-maker. The right vertical axis is for F_∞, the left one is for F_n / F_∞ and $\bar{p}$ (from Arisz, Kanis and Rooden, 2000)*

The example in Figure 2.7 concerns data from a study by Rooden (Rooden et al., 1999): predictions by 20 experts about problems to be expected when using a coffee maker with a wide rage of additional functions such as filtering of the water and a timer setting. Figure 2.7 shows that a reasonable estimation of F_∞ can already be made after five experts, whilst F_n / F_∞ is still only *0.6*, in other words, with 40% of F_∞ estimated at that time still unknown. The fact that F_∞ continues to rise slightly and $\bar{p}$ continues to drop in a similar fashion is not only caused by the averaging of p_i (*i=1 ...* F_n) into $\bar{p}$, but also because more problems with a relatively high p_i will be predicted first by definition; in other words: new problems predicted by the later experts will always cause $\bar{p}$ to drop.

2.8 Analysis

2.8.1 Answering research questions as a guideline



Keeping in mind Figure 2.5, always remember the research questions during the analysis. A productive approach often starts with the usability problems that have been identified, problems encountered by test subjects themselves and/or in your eyes, in terms of unsatisfactory product performance, side effects, extra effort which the user has to put in ('product performance', 'side effects', 'effort' in Figure 2.5). Then it is a matter of understanding the *what* and *how* of what was observed ('use actions' in Figure 2.5) on the basis of the collected video and sound recordings from perception and cognition, among other things, relating to presumed/designed material and functional instructions ('perception', 'cognition' in Figure 2.5), including the role of relevant experience ('experience' in Figure 2.5), whereby human characteristics can also be involved where required ('characteristics', in Figure 2.5).

You should then also consider usage episodes that were successful, especially if a study question has been asked about them, again keeping in mind the representation of the functioning of products in Figure 2.5.

Generally speaking it may be more efficient to work on the basis of (video) stills rather than having to go through the entire tape every time. It is then recommended to compile a separate file of video stills with notable events, possibly sticking them into a log with an explanation of what was the matter with each test subject. Quite often this also appears to be a good individual means of communication. Try to avoid counting at random, such as the number of times a particular finger does something and the time this takes (except of course in the exceptional case that these things are considered relevant and presuppositions have been made and study questions have been drawn up about them). For example, the time people take to do something is just as easy to measure and average in many cases (not always, see the following paragraph) as it is meaningless from a design point of view. Anyone who does start counting to illustrate certain things, for example, attempts, must indicate accurately what distinguishes attempts, when an attempt ends/a new one starts. The following should also be realised here. If it is concluded in the analysis that a test subject looks tense while performing the assignment, this assessment

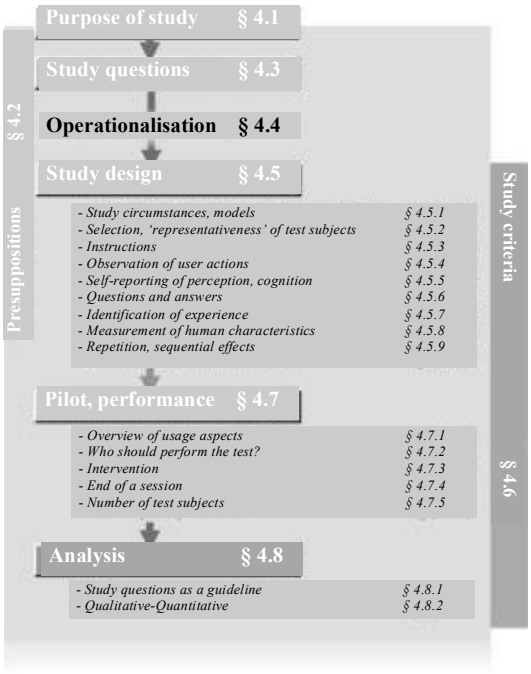
must be reproducible. That is to say, it must be possible to report the criterion of characterising something as tense clearly (rather than stating some kind of vague impression), as a result of which someone else who analyses the visual material will also conclude 'Yes, that is tense'. But that's not all. Tense in whose eyes? Maybe this test subject always does things like this, is acting normally, and you would see something completely different if he or she really had a tense position, in other words, for that test subject personally.

If each test subject has to perform more than one task, check whether there are indications for pass-on effects (see §2.5.9). It is important to be aware of artefacts like these and identify them where possible rather than trying to cover them up. They can be traced in an analysis by comparing the cases with different histories – for example, 'first task a and then task b' and 'first task c and then task b'. This is only possible in a 'within subjects' design with a limited number of tasks and a sufficient number of test subjects, as the number of observations per cell would otherwise be too small.

If you can't find any indications in a case like this relating to effects of different histories, that does not mean that these effects are entirely absent. They were apparently not identified in the observations that were collected as far as you can tell: that is what you can conclude afterwards as a researcher.

2.8.2 Quantitative-Qualitative

Measuring, counting? Imagine you have to hire people for whom quantitative research is all that matters. Others can't stop pointing out that quantitative research can only be carried out once it has been sorted out thoroughly how to categorise, what to count and how. Some people say that, if you know the latter, the problem you wanted to study is practically solved (Labov, 1972). In so far as a sharp distinction can be made between quantitative and qualitative estimations, an essential characteristic of quantitative research is bundling observations into one group, at least classifying them into a limited number of categories. This allows you to measure, by determining the frequency of one or more agreed observation/measuring units. As a rule the information present here is sacrificed to produce averages. Qualitative observation is generally aimed at obtaining insights other than those related to averages and standard deviations. In that case you will initially not have any fixed observation units, as further specification is allowed, up to the extreme case that each observation is unique.



So what does that mean for usage studies? An example. As stated above, usage studies are performed within a design context, initially not with the aim to count the frequency of a particular problem, nor to measure which performance times result in usage problems or not. In usage studies it is first and foremost a matter of identifying the causes of problems as precisely as possible, see Table 2.2. This table shows that observation of user actions as envisaged in a design does not have to mean that cues such as external and functional instructions are noticed and/or understood: it may be a stroke of luck. Similarly, use that differs from what was intended can have many reasons, and definitely not just because something has gone wrong with perception/cognition. As stated before, with regard to design changes to be implemented as a result of problems observed, it makes a huge difference whether users do not observe matters, do not understand them or interpret them differently from what was intended, or simply prefer their own approach. Does this mean that usage studies are only qualitative? No, whenever phenomena are studied that can be quantitatively described in dimensions that are directly relevant to use, in other words, directly related to the material or functional product characteristics to be (re)designed, quantification is the best way to do it. An example of this is the study into foot movements in cars as shown in Figure 2.8. In this study the foot movements of drivers were recorded in such a way that it was possible to quantify them.

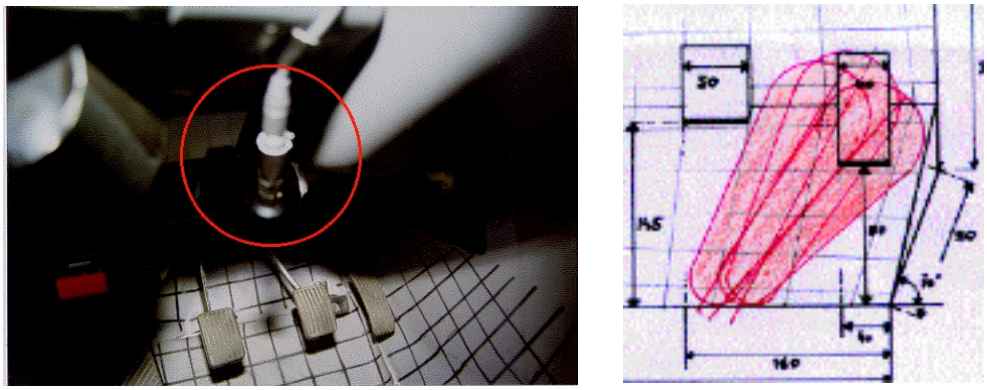


Figure 2.8 *Foot movements when operating pedals in a car (Peijs et al., 1999)*
The circle in the left picture shows the camera. On the right are the furthest positions taken up by the right foot as the throttle was operated during a 55-minute drive (40 minutes in the city, the rest on the motorway).

Another example is the question of whether users tend to press a button for long enough to obtain a particular function that is activated after a certain number of seconds. In this case it would be a huge mistake if you didn't measure the time it takes to perform such a user action. Please note that time is a so-called added dimension in the example mentioned above, whereby the occurrence of user problems as such is converted into spending more time. And what about things like measuring the time spent in a work area? The possible design relevance of such a measure is related to positions taken up and activities performed (and the period involved). As an overall measure it's not much use to a designer. Incidentally, this does not mean that this kind of

quantification is never useful; just think of the significance in terms of business economics of comparing the usefulness of, for example, different office software packages, for example, on the basis of time measurement and/or counting errors.

Please note here that especially behavioural scientists (only psychologists?) regularly refer to the following quote by Kelvin, to prove that quantification is still the only acceptable method to conduct science: ‘When you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind.’ But that is not the way Lord Kelvin, who was a mathematician and physicist (thermodynamics), wrote it. His complete statement was as follows (Sydenham, 1979, p.4): ‘In physical science [...] when you can measure what you are speaking about, and express it in numbers, you know something about it; but [...] when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind [...].’ Kelvin wrote this in 1883, many years before behavioural science evolved into a ‘hard métier’.
The following example should be considered a caricature in this context, also of the approach of even the most experienced *die hard* in behavioural science. In the recent past a group of usage researchers got stuck on the random idea that a test subject who uses a product twice a week belongs in the category ‘highly experienced’. But it didn’t stop there, as they then adopted the foolish idea that the test subject in question would have twice the experience of a test subject who only uses this product once a week.

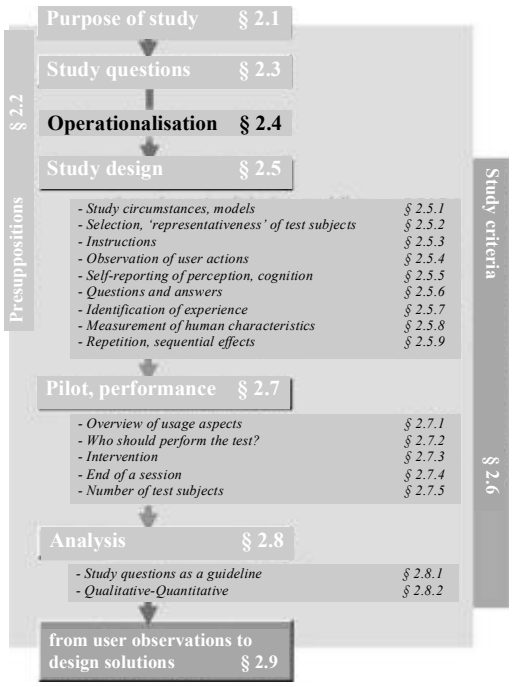
2.9 Towards design solutions, usage observations as a source of inspiration

Imagine finding out in a study that certain user actions with a relatively high risk of accidents are performed to compensate for lack of efficiency. In this case one of the design requirements to be included could be that, after redesigning it, the efficiency of the product should be improved to such an extent that it is reasonable to assume that users will no longer resort to the unacceptable compensating user actions. Another approach in this case could be that these user actions are allowed, but should no longer result in an accident. This means that the substandard functioning can be ignored (whether this is sensible is not the issue here). A third approach is to change the design in such a way that it requires a completely different use. The uncertainty of newly arising problems is, however, greatest in that case.

Which requirements are eventually met depends on the extent to which it is possible to find adequate solutions in each of these cases. Note that the domain of the designer concerns the product: solutions will always have to be implemented via exterior and functional product characteristics. Limits to the possibilities that designers have here are, on the one hand, not imposing any restrictions on user actions and environmental characteristics (users can keep doing what they were doing/are used to, as the problems have been eliminated by changing the product characteristics) and, on the other hand, forcing user actions and environmental

characteristics, see Mauro (1978) who designed a hair dryer that was mounted to the bathroom wall, to prevent a handheld device from falling into the bath and electrocuting bathers.

Within these extremes, observed use with undesirable repercussions for the functioning of a new design can be tackled with adequate material and functional instructions for use. Consider whether there are opportunities to limit the restrictions of a design as much as possible with regard to user actions that result in the intended functionality, see the example of the bicycle parking facility in Figure 2.9. The caption of this figure states an aspect that has not been discussed yet, namely serving as a source of inspiration. See for other examples like this Van Hees et al. (1997) and Klein et al. (1998); both publications are also based on thesis work.



! Always bear in mind that each change for the better in a design introduces uncertainties that may prove to have a detrimental effect. If a usage study proves that a number of presuppositions were wrong, it will not look smart if you think you can praise proposed changes without basing yourself on a new usage study. In a case like this it seems that little has been learned. And if a usage study is concluded without any problems, please don't think you can come up with the conclusion that the product or model is perfect in every way. What's the main point of a usage study again? And what's the flaw of not having enough test subjects? And what would have been the result if the devil himself rather than his advocate had used the model/prototype/product?

2.10 About reporting: suggestions, possible points for attention

☞

Not everything below will be equally applicable all the time. Make sure that, by sticking to essential matters, the study report is no thicker than this manual. Never repeat this manual for explanations, substantiating, clarifications. Refer to it, that is why this manual is as thick as it is.

Here are some suggestions and possible points for attention:

- Where required help readers on their way with a clear picture/presentation drawing of the artefact (model/prototype/working product), stating the most essential designed/assumed material and functional instructions or usecues that are discussed. Note: don't make up things that assign a deeper meaning to every curve and colour difference afterwards, in other words, don't 'usecue the matter to death', see the end of §2.2. Don't include any technical drawings. Include material and functional instructions in pictures, enlarged where necessary;

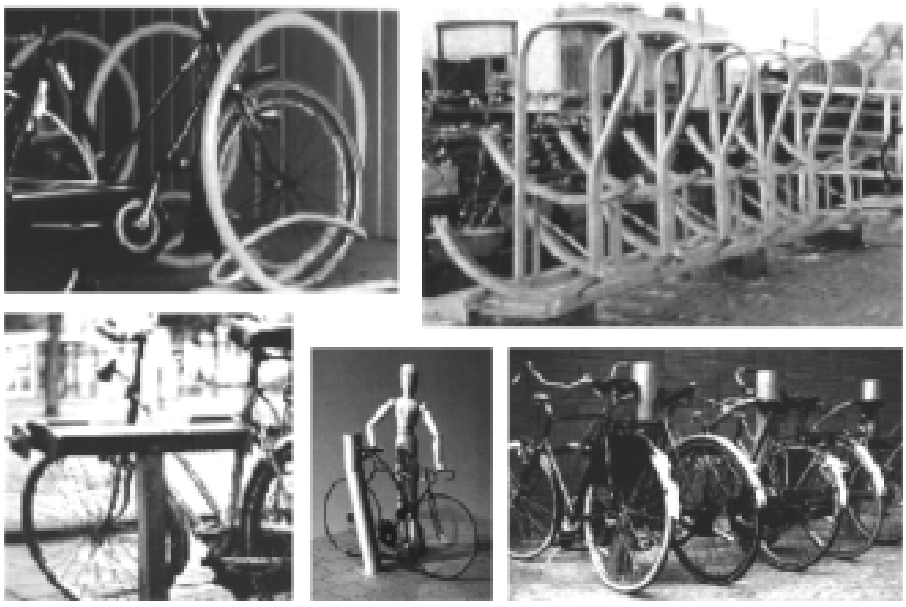


Figure 2.9 *A new bicycle parking facility (Van der Steen, Kanis, Marinissen, 1996, graduation project for Amsterdam city council)*
A usage study showed that most of the numerous usage problems faced with various existing racks (top row) were a direct result of the precise use required as recorded in the various designs. In other words: the existing racks had to be used in a particular way for the designed functions to be as effective as possible, for example, the possibility of storing a bicycle in a well-supported manner and locking it. After that this insight did not lead to some kind of redesign, but a completely new concept was developed (bottom row), among other things, inspired by the way in which you can see many bicycles parked in Amsterdam, for example, chained to lampposts and bridge railings. The new design targeted a lot of freedom of use without losing any essential functionality.

- use a limited number of words in any case: a description of visual cues in words is obviously unacceptable. Usecues/material and functional instructions – or whatever you would like to call them – can, for example, be numbered in view of references in the text.
- A storyboard of the use can often be enlightening for readers, for a fast and adequate insight into the order of possible actions. To avoid any misunderstanding: the use does not have to be the same as presented in the storyboard, as that is precisely what the study is for.
- The presentation picture/drawing of the product and the storyboard of the use can be combined into a single storyboard, where required, with text.
- Describe presuppositions (§2.2) in one or two sentences per material and functional instruction; describe *why* they work (or don't work on reflection). Be careful not to assume a study is 'dead', see the end of §2.2. If it is impossible to indicate a deeper reason for certain aspects of the acts and omissions of users, state that it appears to be self-explanatory, literally right in front of you, plausible. Where applicable, state why human characteristics and capacities play a role. Should the study happen to yield clues that, for example, older people act differently than the rest after all, this should be investigated further, as age itself (a *proxy*,

see §2.3) does not provide any clues for a designer to intervene in material or functional product characteristics; the same applies to demographic characteristics such as gender and occupation.

- Phrase research questions (§2.3) accurately. Originality does not have priority; the fact that it is well-considered must come first. Make sure that the order is logical, as the presentation of the findings consists of answering the study questions, see further down.
- Give a brief description of the research design, including the pilot (§2.5.7). Have any new presuppositions been included inadvertently as the study circumstances were put into practice? Did test subjects receive things in the same manner for every (sub)assignment (§2.5.3), in other words, did the model/prototype/working product have the same setting, additional materials in the same location, etc.? Were efforts made to achieve unobtrusive observation (§2.5.1, §2.5.3)?
- Instructions (§2.5.3) should be written in colloquial language, straight to the point without wasting words, as simply as possible and preferably only in terms of an end result to be achieved without describing all the steps in between.
- Overview of points for attention relating to use as a specification of the study questions, don't include all sorts of information (§2.7.1) in this overview on which test subjects have to comment. At least include the version that was used at the start of the study, for the first test subject, as an appendix (a list of points for attention relating to use like this will often get longer during the study as a result of unexpected things that are observed).
- Make a checklist of matters that have to be sorted out when a new test subject enters, to make sure that everything is ready, working and you can't miss anything.
- What was learned from the pilot, what was the further progress of the study, what went right in terms of study performance, what went wrong.
- It may greatly help readers if, before the results are presented, a brief description is given of the anonymised test subjects in terms of their characteristics, in so far as these are mentioned in the study questions (as a rule nothing about gender, age, occupation). The reason for having this description precede the answers to the study questions is the following. There is always one study question about the role of individual characteristics and qualities: are the actions and omissions of test subjects linked to certain experience, capacities (eyesight, manual power). These questions can be discussed in a way that provides readers with an insight if it is possible for cases observed to refer to test subjects involved on the basis of their personal characteristics and qualities: experience, capacities, anthropometric characteristics in so far as they apply. Test subjects will come to life. This description of test subjects in terms of characteristics should preferably be presented in a diagram for a quick overview.
- Report the answers to the study questions 'on a level beyond test subjects', not for each test subject (experiences per test subject can be added as a appendix). The study questions are the basic structure, to which basically all findings can be traced back. If events were observed that are worth mentioning but cannot be classified properly under any of the study

questions, these should be amended. Furthermore, an illogical order of the study questions to be answered may lead to the findings being reported in a highly non-transparent manner. Imagine: first answering the question *Which of the presuppositions are correct?* (assuming it has been asked) without you as a reader having any idea of what happened, and then the question *What are the actions and omissions of users and what problems do they have?* (again assuming that this type of question has been asked) Don't think that this is an invented example: it really happened.

If you have a study theme, ensure that the answer to the relevant study question is as detailed as possible.

- Pictures/images should always be included in the main text. If a picture is only good enough for an appendix, the question arises whether a picture like that would add anything, whether it shouldn't be removed completely.
- Raw data, that is to say, the data as collected for each test subject – think about notes that were recorded – must not be reported, not in the appendices and absolutely not in the main text.
- Stay close to the data. Don't be afraid to quote test subjects if it concerns concise, clarifying statements. Quote test subjects literally, don't 'improve' their language. A study report definitely does not have to turn into a regional novel.
- Correct language is appreciated. And now that we are talking about it: don't write '... there were 8 test subjects in total. 2 of them...'; write 'There were eight test subjects in total. Two of them...'.
- Are there clues for pass-on effects (§2.5.9, only for more than one task)?
- It often happens that an exhaustive overview is given of the usecues that are considered essential, often in a diagram, stating for each material and functional instruction whether test subjects have observed them or not, and if they have understood them and acted upon them. An overview like this is absolutely not a 'must', usecues are never an end in themselves, only a means. Additionally, such a diagram often gives rise to the suggestion that each material and functional instruction would work by itself, not in context. A diagram like this also acts as an invitation to fill it in completely, as if all test subjects have been interviewed about all material and functional instructions in terms of their perception and cognition. Is this true? A fully completed diagram without any trace of doubt does not look reliable. It's better to try and say something about clues as to the contextuality of cues. And you could, for example, check which of the cases stated in Table 4.1 ('User activities for various reasons other than those (supposedly) intended in the design') have been observed. And wouldn't there be observations for which this table isn't suitable?

  If you include a diagram – have designed/presumed usecues that are considered essential been noticed/understood/acted upon? – always summarise it in terms of successful or failed mediation of usecues, also to prevent reporting from being only test subject based and to prevent the readers from having to figure it out themselves. And again: figuring out the functioning of designed/presumed usecues – or whatever you would like to call them – is

never an aim in itself, in the end it's always about the mediating role of material and functional instructions in whether or not any usage problems arise.

- Has anything (else) unusual been observed? For example, test subjects for whom it all went smoothly, but who appeared to have failed to notice or understand things upon further questioning, for example, afterwards. Are there any test subjects who deny afterwards they have done something that was recorded on video, see examples in §2.5.6 (*Questions & Answers*)?
- Usage studies performed in a scientifically oriented environment require proper recording, incidentally, with every freedom to refer to this manual to avoid saying things twice, filler text, see the beginning of this paragraph. New designs are generally created within an applied environment. There is no time there for exhaustive reporting. Even more so, long-winded stories are read least often, they have the greatest chance of ending up in a drawer straight away or worse. Rather than summarising results on an A4 sheet, it is better (and unfortunately also much more time-consuming) to make a video compilation, allowing everyone to convince themselves (or be convinced), up to and including the management. Do remember that the fiction of objectivity also applies to a summary like that on video: it consists of your choices. Nevertheless, one of the most important guidelines for compiling a video compilation of a limited number of minutes is to speed things up. You can always stop the video for details.