Image Schemas and Design for Intuitive Use

Exploring New Guidance for User Interface Design

vorgelegt von
Diplom-Psychologe
Jörn Hurtienne

Von der Fakultät V – Verkehrs- und Maschinensysteme
der Technischen Universität Berlin
zur Erlangung des akademischen Grades
Doktor der Ingenieurwissenschaften
- Dr.-Ing. -

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender:	Prof. Dr.-Ing. Matthias Rötting
Gutachter:	Prof. Dr. Phil. Manfred Thüring
Gutachterin:	Prof. Dr.-Ing. Luciënne Blessing

Tag der wissenschaftlichen Aussprache: 10.12.2009

Berlin 2011
D 83

Abstract

In this work, intuitive use is defined as the extent to which a product can be used by subconsciously applying prior knowledge, resulting in an effective and satisfying interaction using a minimum of cognitive resources. Image schemas are proposed as new design guidance, and the usefulness of image schema theory for design for intuitive use is investigated.

Image schemas are sensorimotor and subconscious forms of knowledge representation. Thus, they fulfil the preconditions of intuitive use and hold great promise for user interface design. The promises of image schema theory are discussed in the light of existing linguistic and psychological research and empirical research questions are derived from this discussion.

The first research question relates to using image schemas for conveying abstract information in user interfaces. The results of four experiments show that user interaction can be more effective, mentally efficient, and satisfying with user interfaces that conform to image schema theory than with user interfaces that do not conform to the theory. The size of the effect is dependent on the task given to the users, the difficulty of the task, and the presence of other image schema instances in the user interface.

The second research question relates to the practicability of image schemas as a design language for designing intuitive use. Two studies investigate the inter-rater-reliabilities when applying an image schema vocabulary to the description of tasks, interactions, user interfaces, and user utterances. The results show high to medium agreement among the designers using the image schema vocabulary. In another study, designers used the image schema language in a human-centred design process for the development of two new prototypes of an existing enterprise resource planning system. Image schemas proved especially useful in the process of translating requirements into design solutions. To support future design processes, an online database was developed that gives product designers access to image schema definitions and provides examples of image schema instantiations in user interfaces.

As a result of these studies it can be concluded that image schema theory provides valid heuristics for designing intuitive use and that applying an image schema design language is reliable and practical and can benefit the early phases of product design. Image schemas not only complement but even go beyond the scope of the existing guidance for designing intuitive use like user interface metaphor, population stereotypes, or affordances. Questions for further research are discussed.

Kurzfassung

‚Intuitive Benutzung' wird in dieser Arbeit als die Benutzung eines Produktes definiert, die in unterschiedlichem Maße durch die unbewusste Anwendung von Vorwissen charakterisiert ist und zu einer effektiven und zufriedenstellenden Interaktion bei minimalem Verbrauch kognitiver Ressourcen führt. Als neues Gestaltungsmittel mit hohem Potential werden Image Schemata vorgeschlagen und die Nützlichkeit von Image Schema Theorie für die Gestaltung intuitiver Benutzung wird untersucht.

Image Schemata erfüllen als sensumotorische Form unterbewusster Wissensrepräsentation die Voraussetzungen für intuitive Benutzung und ihr Einsatz in der User-Interface-Gestaltung ist vielversprechend. In Hinblick auf bereits vorhandene Forschungsarbeiten in der Linguistik und Psychologie wird das Potential der Image-Schema-Theorie diskutiert und es werden empirische Forschungsfragen abgeleitet.

Die erste Forschungsfrage betrifft die Anwendung von Image Schemata bei der Darstellung abstrakter Informationen in User Interfaces. In vier Experimenten wird gezeigt, dass Benutzer effektiver, mental effizienter und zufriedener mit theorie-konformen User Interfaces interagieren können als mit User Interfaces, die nicht theorie-konform gestaltet sind. Die Größe des Effekts ist dabei abhängig von der konkreten Aufgabe, der Aufgabenschwierigkeit und dem Vorhandensein weiterer Image-Schema-Instanzen im User Interface.

Die zweite Forschungsfrage betrifft die praktische Einsetzbarkeit von Image Schemata als Gestaltungssprache bei der Entwicklung intuitiv benutzbarer User Interfaces. In zwei Studien werden die Inter-Rater-Reliabilitäten bei der Anwendung eines Image-Schema-Vokabulars untersucht, wobei sich zwischen Designern hohe bis mittlere Übereinstimmungen bei der image-schematischen Beschreibung von Aufgaben, Interaktionen, User Interfaces und Benutzeräußerungen ergeben. In einer weiteren Studie wandten Designer das Image-Schema-Vokabular in einem nutzerzentrierten Gestaltungsprozess an und entwickelten zwei neue Prototypen eines bestehenden Warenwirtschaftssystems. Dabei erwiesen sich Image Schemata als besonders nützlich bei der Umsetzung von Anforderungen in Gestaltungslösungen. Die image-schematisch gestalteten Prototypen wurden von denn Benutzern besser beurteilt als das bestehende System. Zur Unterstützung von Entwicklungsprozessen wurde eine Online-Datenbank entwickelt, die Produktentwicklern Definitionen von Image Schemata und Beispiele für ihre Anwendung in User Interfaces zugänglich macht.

Als Ergebnis der Studien lässt sich feststellen, dass die Image-Schema-Theorie gültige Heuristiken für die Gestaltung intuitiver Benutzung liefert. Eine Image-Schema-Gestaltungssprache ist reliabel und praktisch anwendbar und kann in frühen Phasen der Produktentwicklung nutzbringend eingesetzt werden. Bisherige

Ansätze zur Gestaltung intuitiv benutzbarer Produkte wie User Interface Metaphern, Populationsstereotypen oder Affordances können durch Image Schemata nicht nur ergänzt, sondern zum Teil in der Breite ihrer Anwendung übertroffen werden. Offene Fragen für die weitere Forschung werden ebenfalls diskutiert.

Acknowledgements

This work is the result of my research at the interdisciplinary research group prometei (prospective design of human-technology interaction) and the Chair of Engineering Design and Methodology at Technische Universität Berlin. Many people within and outside these groups have contributed to the process of the research that led to this work.

I need to express my most special thanks to Lucienne Blessing and Manfred Thüring who helped me to take on the interdisciplinary challenge of this work – bridging psychology and engineering by a linguistic theory. They greatly supported me with advice, with practical problem solving, and by challenging the ideas I developed during this research. Although their own research agendas were different from mine, they were always greatly supportive, full of interest for my work, and open for discussion at any time it was needed.

For the original inspiration for this research topic I highly appreciate the input of Juliana Goschler who as a linguist introduced me to conceptual metaphor theory, which immediately sparked the idea of trying it in user interface design.

The studies presented here owe much to the help of my students and colleagues. Maria Luz and Kathrin Loest helped with gathering the data in the experiments. Luisa Hahn and Su Dong contributed much to the content of the ISCAT image schema database, and I am grateful to Jan Groth for helping with the technical realisation. A big thank you goes to the H-metaphor group at the DLR in Braunschweig, especially Frank Flemisch, Anna Schieben, and Johann Kelsch, for the vivid discussions and their continued interest in FORCE image schemas. Big thanks go to Katharina Weber and to Johann Habakuk Israel who as designers helped with exploring the application of image schemas as a prescriptive language for design.

Many thanks need to go to the colleagues and administrators of the prometei research group, of the Engineering Design and Methodology group, especially Michael Schmidt-Kretschmer, and the researchers associated with the Chair of Cognitive Psychology and Cognitive Ergonomics (all TU Berlin) for their continuous feedback and their general support. For enlightening discussions of intuitive use I would like to thank the IUUI research group (Intuitive Use of User Interfaces), especially Carsten Mohs, Habakuk Israel, and Herbert A. Meyer.

Writing it up is the most arduous task and many friends and colleagues gave invaluable support. I am especially grateful to Anna Pohlmeyer, Habakuk Israel, Karen Krüger, and Carmen Bruder who struggled through the first versions of this work and gave precious feedback on enhancing it. Many thanks go to Petra Abele, who took on the role of the reader's advocate in later phases, reminding me of the red thread and better structure. For proofreading the final version I thank my partner Martina Englert, who helped me to maintain proper English and a good spirit during those final days and late nights. She also provided continuous and loving support during the whole time of thesis research and writing.

Table of Contents

List of Abbreviations

2D	Two-Dimensional
3D	Three-Dimensional
ANOVA	ANalysis Of VAriance (statistical procedure)
CAD	Computer Aided Design
CD	Compact Disc (technology)
d	Symbol for effect size (statistical parameter)
df	degrees of freedom (statistical parameter)
ERP	Enterprise Resource Planning
GLM	General Linear Model (statistical procedure)
GUI	Graphical User Interface
HTI	Human-Technology Interaction
IS	Image Schema
ISCAT	Image Schema Catalogue (database)
ISO	International Standardisation Organization
IUUI	Intuitive Use of User Interfaces (research group)
LCD	Liquid Crystal Display (technology)
LED	Light-Emitting Diode (technology)
M	Mean (statistical parameter)
MABA-MABA	Men Are Better At – Machines Are Better At (Fitts, 1951)
OCR	Optical Character Recognition (technology)
p	probability (measure of statistical significance)
PC	Personal Computer
PDA	Personal Digital Assistant (mobile device)
PIBA-DIBA	Physicality Is Better At – Digitality Is Better At (Hurtienne, Israel, & Weber, 2008)
PMA	Perceptual Meaning Analysis (Mandler, 1992, 2004, 2005)
RFID	Radio Frequency Identification (technology)
RR	Representational Redescription (Karmiloff-Smith, 1995)
RSI	Repetitive Strain Injury
SAS	Supervisory Attentional System (Cooper, Shallice, & Farringdon, 1995)
SD	Standard Deviation (statistical parameter)
SNARC	Spatial Numerical Association of Response Code (e.g. Gevers, Reynvoet, & Fias, 2003)
TCO	Total Cost of Ownership
TUI	Tangible User Interface
UCD	User Centred Design
UI	User Interface
UID	User Interface Design
UIM	User Interface Metaphor
WWW	World Wide Web

1 The Demand for Intuitive Use

Intuitive use currently is a buzzword. Products have *intuitive interfaces* (e.g. Briell Marketing, 2008; ControlGlobal, 2008; Hurricane Software Inc., 2002), research labs include *intuitive usability* in their research agendas (Deutsche Telekom, 2007), patents claim *intuitive controls* (Hadank, Allen, Bradbury, & Anderson, 1991; Karem, 2001), and the European Union aims at *developing user friendly interfaces which are intuitive* (IST, 2006).

So why is it that *intuitive use* is sought after while the seemingly similar concept of *usability* has been around for years? Intuitive use can be seen as a sub-concept of usability with a strong focus on the mental demands in using technology. As these mental demands grow, the need for intuitive use grows accordingly. Several trends contribute to this.

One trend is the increasing number of functions in products. Mobile phones, for example, have been equipped with cameras, media players, games and time-planning software. At the same time, the functionality becomes more sophisticated: the camera in the mobile phone offers a choice between a night and a day mode, between macro, portrait, and landscape mode and between a range of flash modes. The increasing functional complexity is an issue for design – as witnessed by the titles of recent best-sellers in technology design: *Don't Make Me Think!* (Krug, 2006) or *The Laws of Simplicity* (Maeda, 2006).

A second trend is subtler: hardware is turning into software. Human-machine interaction becomes human-computer interaction. The control room of a power plant or the operating controls of a break press, once designed as hardware, have turned into software user interfaces. The personal computer (PC) is an extreme example. Its hardware is reduced to the standard components of a mouse, a keyboard, and a monitor screen, while its functionality is delivered by software. The flexibility of software allows the user interface design to be decoupled from the underlying technology – a chance for designers that rarely is made use of to the full extent. The disadvantage of the PC is that key-presses and mouse movements only enable low-bandwidth interaction compared to the capabilities of the human sensorimotor system (Ishii & Ullmer, 1997).

A third trend is that technology moves away from controlling physical things, e.g. steering cars or controlling chemical processes. New fields of application involve the interaction with abstract data, e.g. managing financial data in enterprise resource planning systems, updating a profile on a dating website, or caring for the health of a virtual pet. Although the content manipulated with the help of technology is becoming abstract, input and output devices remain primarily spatial in nature. How can spatial coordinates of an input device be related, or mapped, to abstract data? A major task for developers is to make such mappings intuitive.

A fourth trend goes toward more heterogeneous user groups. For consumer goods, the targeted user groups are highly variable in age, experience, and cultural background. Virtually everyone should be able to use a ticket or cash machine, an interactive TV set, or an in-car driver assistance system. A user interface that does not accommodate the different backgrounds of users will not be intuitive to use.

These problems get worse by the sheer number of interactive products. Too many different devices accrue in the users' environment, so that the time and motivation available for learning and using each device is greatly reduced. Providing documentation or training cannot counterbalance bad user interface design. Users do not read manuals (Marc, 1991) and are not likely to attend training classes to learn about their devices. Again, if devices were more intuitive to use, the need for training would be reduced.

An important factor for 'intuitive use' becoming a buzzword certainly is marketing. As the technology and functionality of products in the same market tend to converge, intuitive use becomes a unique selling proposition**.** Particularly for investment goods, the total cost of ownership (TCO) is an important issue. Traditionally this involved the costs of a products' energy consumption, its maintenance, and disposal. Only recently, the productivity of the users interacting with the product also became an essential issue for designing machine tools, business software, or smart phones (e.g. Abele, Hurtienne, & Prümper, 2007; Zühlke, 2004). As time is money, user interfaces that are too complicated to use lower the productivity of the workforce. User interfaces that are intuitive to use reduce training and support costs – a fact that is very important in departments with high employee turnover rates (e.g. call-centres). Finally, safety and security issues are paramount for driving vehicles, flying planes, operating power plants, performing computer assisted surgery, etc. In these domains human error rates need to be kept infinitely small. Technology that is intuitive to use can contribute to this.

To summarise, increasing demand for intuitive interaction arises as complexity grows, more hardware turns into software, new fields of applications and new user groups emerge, and the number of interactive products increases. An increasing dependency on technology will fuel the demand for intuitive use at the workplace, at home, in e-learning, in e-commerce, in e-government, in entertainment, and many other areas.

This demand for intuitive use brings about several requirements that are addressed by the methods and principles discussed in this work. The main issue here is how designers can design for intuitive use. In the following subchapter, a principle is introduced that is the foundation of successful user interface design.

1.1 A Basic Principle for Designing Intuitive Use

In modern technology, user interfaces can be decoupled from the underlying technology. It has become common in user interface design to distinguish between the implementation model, the represented model, and the mental model of the user (Cooper & Reimann, 2003; for a similar discussion see Norman, 1990). The implementation model is the model of the technical system, of its working principle, the mechanics, or in the case of software, the algorithms or the database (*Figure 1.1*).

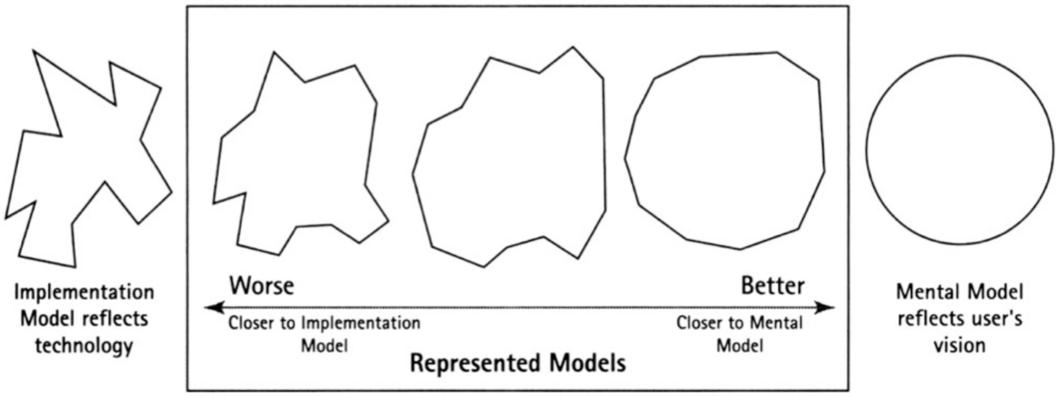


Figure 1.1. Good user interface design matches the mental model of the user (from Cooper & Reimann, 2003, p. 23)

On the other side, there is the user's mental model. This represents what the user knows of the workings of the system and of the purposes he or she wants to achieve by using it. Mental models do neither contain all the details of how a product works nor do they correctly represent the underlying implementation model (Norman, 1983). They are used as a cognitive shorthand covering purposeful interactions. In a telephone, for instance, users expect the same behaviour regardless whether it is a mobile or a landline phone. They do so, although the two types are based on very different implementation models, i.e. principles of transmission. In software, the decoupling of the mental model from the implementation model can be particularly strong. The users' mental model of using the Google search engine is simple. Its underlying implementation model is tremendously complex – involving communication protocols, databases and retrieval algorithms.

By designing the user interface, designers make decisions on how the product represents its functioning. That is, they determine how the user interface looks, sounds, feels, and behaves. This representation of the technology in the user interface is captured in the represented model (*Figure 1.1*).

The represented model can be chosen to be close to the true actions of the product determined by algorithms and mechanical constraints. Alternatively, the designer can choose to hide much of the inner workings of the technology from the user. The closer the represented model is to the user's mental model, the easier the user will find the product to use and understand. As the user's mental model of his task is bound to differ from the inner workings of the product, a represented model that is close to the implementation model may burden the user with the task of learning and re-learning to operate the product each time he uses it. Represented models that are close to implementation models are therefore less likely to be intuitive to use.

User interface designers have more control over the represented model than over the other two models. Therefore, they should try to design the represented model to be as close as possible to the mental model of the user. Although this may sound trivial, too many user interfaces do not follow this philosophy. A book-keeping application, for example, mirrors the structure of its database tables on the user interface rather than the steps needed for posting invoices. A computer-aided design (CAD) application requires the designer to enter exact numbers when specifying the geometry of objects, although the user only wants a quick sketch for visualising an idea.

Even if they are formally committed to intuitive interaction, designers can make the wrong decisions. In a misunderstood sense of simplicity, the designers of the BMW iDrive provided just one control knob and one display unit for accessing over 700 settings controlling the navigation system, the audio system (radio and CD player), the phone, and the climate control unit (Norman, 2004). With this design, the iDrive mimics the control options of a personal computer – a narrow bandwidth human-technology communication channel that accesses a highly flexible software interface. Although the *i* in iDrive stands for *intuitive*, iDrive has caused significant controversy among users. Criticisms include long navigation paths to access simple functions, slow feedback, cryptic menu labels, a steep learning curve, and a tendency to cause the driver to look away from the road too much (Cobb, 2002; Kujik, 2007).

Of course, if the users need to know about or are interested in the actual workings of the product, then the product can profit from a transparent implementation model. In investment goods it is sometimes necessary to understand the technology behind the user interface (e.g. in process control or aviation). In consumer products this might enlighten technology enthusiasts, but it will not make a user interface intuitive to use for most of the other users.

So what is going wrong, when even products that are labelled *intuitive to use* do not live up to their promise? One reason is the different understandings of intuitive use people have. Although intuitive use is demanded by users and frequently advertised by producers, they rarely make explicit what intuitive use is. Only by defining the concept and communicating it, designers can develop a common understanding of what to design for. A second reason could be that, although the goal is clear, there are no ways of measuring goal achievement, i.e. means of measuring intuitive use. It needs to be clear how intuitive use can be operationalised and how it can be measured. Otherwise designers will not know whether they have reached their goal or not.

Third, and maybe most importantly, designers might lack appropriate guidance on how to design for intuitive use. It still seems that many user interface designs are technologically driven. New technologies open up new interaction possibilities and these are applied to almost any task domain. The basic principle of design suggests otherwise. Design should be driven by the tasks of the users and the users' mental models of these tasks. User interfaces should be structured in the same way as these mental models, and the technological implementation needs to follow. Although several approaches exist to discern the contents and the structure of tasks and users' mental models, they seem to be of limited scope, usefulness, or practicability. However, even if there are good tools for the analysis of tasks and mental models, this does not mean that the results easily translate into a prescription on how to match the represented model to the mental model of the user.

This work provides approaches to user interface design that can help designers build products that are intuitive to use. To this end, a definition of intuitive use is provided, possible ways of measuring intuitive use are discussed, and a brief review is given of the guidance that is currently available to designers for analysing the mental models of users and matching representation models to these mental models. It is acknowledged that the existing design guidance has limitations in terms of scope, theoretical foundation and its ability to meet the demand for intuitive use as sketched in the beginning of this chapter.

Therefore, in this work the search space is widened to include concepts not traditionally discussed in the field of human-technology interaction. The focus lies on image-schema theory originating in linguistics and philosophy (Hampe & Grady, 2005; Johnson, 1987). Image-schema theory promises to deliver useful guidance for describing users' tasks and mental models and prescribing user interface designs, i.e. representational models. The theory also promises to be useful in designing for heterogeneous user groups and the manipulation of abstract data, and can be applied to the design of hardware and software user interfaces alike. In this work it is investigated whether applying image-schema theory to user interface design can substantiate these promises and support designing products that are intuitive to use. Because of its central role in this work, image-schema theory is introduced in the following section.

1.2 Image Schemas – Promises for Designing Intuitive Use

Image schemas are assumed to be a subconscious form of human knowledge representation that is derived from basic sensorimotor experiences. More specifically, they are abstract representations of recurring dynamic patterns of bodily interactions with the environment: movements of the body, manipulation of objects, and perception. Without these patterns "our experience would be chaotic and incomprehensible" (Johnson, 1987, p. xix). A prominent image schema is the UP-DOWN image schema that forms the basis of

> *thousands of perceptions and activities we experience every day, such as perceiving a tree, our felt sense of standing upright, the activity of climbing stairs, forming a mental image of a flagpole, measuring the children's heights, and experiencing the level of water rising in the bathtub.* (Johnson, 1987, p. xiv)

This example shows that image schemas are not reduced to the visual domain. They can also be derived from acoustic, haptic, and kinaesthetic experiences. Image schemas are much more abstract than mental images and they are represented in analogical form. The basis of our daily experiences with houses, rooms, boxes, teapots, cups, and cars, for example, is captured in the abstract CONTAINER image schema. A CONTAINER is characterised by an inside, an outside, and a boundary between them.

As another example, it is easy to form a mental image of an egg timer in form of a sand glass. The image schemas involved are even more basic: the sand glass is an OBJECT that consists of two CONTAINER image schemas connected via a LINK. Both CONTAINERS are arranged above each other (UP-DOWN). Each CONTAINER can have different degrees (SCALE) of being FULL with sand (SUBSTANCE) or being EMPTY.

Image schemas originate in philosophical and linguistic analyses (Hampe, 2005; Johnson, 1987; Lakoff, 1987) and have been confirmed by recent research in the cognitive (neuro-)sciences (Gibbs, 2005; Gibbs & Colston, 1995; Rohrer, 2005). About 40 of such image schemas can be distinguished (cf. Table 4.1, chapter 4). Image schemas are thought to be universal and underlying more complex mental representations (cf. chapter 3).

Although image schemas are derived from our experiences with the physical world, the human mind re-uses them for structuring abstract concepts (Johnson, 1987; Lakoff & Johnson, 1980, 1999). This is motivated by co-occurrences of image schemas with other experiences. For instance, filling and emptying containers or adding and taking sheets of paper from piles on one's desk lets us experience the correlation of UP-DOWN with quantity. *More* is associated with UP, *less* with DOWN. These correlations of quantity with level of height are so frequent in daily life that they become deeply engrained in our memories. The basic knowledge of MORE IS UP – LESS IS DOWN, is automatically reused when thinking about abstract quantities that are no longer correlated with physical UP-DOWN. Hence, MORE IS UP – LESS IS DOWN has become what is called a metaphorical extension of the UP-DOWN image schema to the domain of quantity. This metaphorical extension can be found, for example, in stock market charts and in verbal expressions like *my income rose last year* or *stock prices fell by 2%.*

With regard to their status as mental representations, image schemas can be instantiated in the represented model, the user interface. They are interesting for designing intuitive interaction in two ways. First, they can be used to instantiate simple physical-to-physical mappings in user interfaces. A LEFT-RIGHT image schema, for example (along with an UP-DOWN image schema) may be instantiated by a mini joystick on a toy car's remote control. When the joystick is moved leftwards, the toy car turns left. A rightward move with the joystick lets the toy car turn right. Second, image schemas can be used in physical-to-abstract mappings using their metaphorical extensions. This way, designers are able to map physical and spatial properties of input and output devices to the abstract concepts users have of their tasks. The metaphorical extension MORE IS UP – LESS IS DOWN, for example, is used to map vertical slider movements to control the volume of the speakers on a PC.

In summary, basic recurring multimodal (visual, acoustic, haptic, kinaesthetic) experiences lead to image schemas. Image schemas represent these experiences in an abstract analogue form. The co-occurrence of such image schemas with other experiences leads to metaphorical extensions, which structure abstract concepts. Image schemas and metaphorical extensions can be used to design user interfaces. To do so, it is necessary to instantiate them in the user interface, i.e. to represent them via the location, appearance, or behaviour of user interface elements.

Image schemas hold several promises for designing intuitive use, which are:

1. Image schemas are very abstract and multimodal structures of knowledge. They could be used to capture user's mental models and reflect them in the design of graphical and multimodal user interfaces. Hence, image schemas could be used as the vocabulary of a metalanguage for user interface design.

2. Image schemas derive from sensorimotor experience. As basic sensorimotor experiences are shared by many users (also across cultures), image schemas should be ubiquitous. In this case, user interface designs created with image schemas should be able to accommodate for heterogeneous user groups.

3. Image schemas are multimodal and can be instantiated in different ways. This suggests that they are applicable to hardware and software user interfaces alike, making them suitable for the trend towards software user interfaces.

4. The metaphorical extensions of image schemas to structure abstract concepts allow designers to convey abstract meaning in the user interface using the available spatial and physical means for input and output. Thus, the trend towards more abstract data represented in the user interface could be met.

5. Image schemas are, in the course of life, extremely frequently encoded in and retrieved from memory. They therefore are supposed to be processed automatically and subconsciously. These processing characteristics are prerequisites for high mental efficiency and for achieving intuitive use.

These promises are taken up in this work, with a focus on promises 1, 3, 4, and 5. The scope of the work is derived from the demand for intuitive use and these promises, as described in the next section.

1.3 Scope

The goal of this work is to determine whether image schemas can be a suitable guidance for designing intuitive use. To this end the following preconditions must be met:

- Because intuitive use is the design goal, an explicit definition of intuitive use needs to be determined. To investigate the extent to which the design goal is met, intuitive use needs to be operationalised, and thereby made measurable.
- Because image schemas are claimed to be a subconscious form of knowledge representation and this claim so far rests on the analysis of phenomenological and linguistic data, the psychological reality of these claims needs to be reviewed in the light of previous empirical evidence.
- Because image schemas, in theory, promise to meet several of the demands for intuitive use, the existing evidence for these promises needs to be reviewed.

After these preconditions are established, empirical studies are derived that address research gaps in the available evidence and attempt to substantiate the promises of image-schema theory for user interface design. More specifically, the following research goals are derived:

- To evaluate physical-to-abstract mappings in user interface design and their contribution to intuitive use. This contributes to investigating promises 4 and 5 above.
- To investigate the reliability of an image-schema design language. i.e. establishing whether designers agree on how they apply the image-schema vocabulary to describe usage scenarios, user utterances, or user interface elements. This contributes to investigating promise 1 above.
- To explore the practicability of an image-schema language in a user interface design process. This contributes to investigating promises 1 and 3 above.

Once these goals are met, the foundation is laid for investigating the value of image schemas in designing for large and heterogeneous user groups (promise 2). This topic, however, is left to further research.

From the assessment of the preconditions for using image schemas in user interface design and the empirical studies a preliminary evaluation is derived regarding the potential of image schemas as a tool for designing intuitive use. A summary of how image schemas compare to the current design guidance is given and a recommendation whether the image-schema approach should be further pursued in the theory and practise of designing intuitive use can be derived.

1.4 Overview

Chapter 2 reviews the research on intuitive use. A definition of intuitive use is provided and the relationship between intuitive use and the concept of usability is discussed. A brief overview is given on what tools have been proposed in the literature to help designing intuitive interaction.

Chapter 3 introduces image-schema theory. It also reviews the available evidence on the suitability of image schemas to meet the defined preconditions of intuitive use and the demand for intuitive use.

Chapter 4 introduces the set of image schemas that is used as the vocabulary of the design language in further studies. It also identifies open research questions and provides a review of previous applications of image schemas in user interface design. Finally, it determines the scope of the empirical studies of this work.

The next two chapters experimentally evaluate hypotheses about metaphorical extensions of specific image schemas. Chapter 5 presents two studies (1 and 2) involving the UP-DOWN and LEFT-RIGHT image schemas. Chapter 6 presents two studies (3 and 4) on contrasting metaphors of the NEAR-FAR image schema. In a review of all four experiments it is discussed whether hypothetical physical-to-abstract mappings can be confirmed in a user interface design context.

Chapter 7 focuses on the application of image schemas as a descriptive language. In studies 5 and 6, inter-rater reliabilities are estimated for assigning image schemas to usage scenarios, user utterances, and user interface elements. In a summary the practical implications of these findings are discussed.

Chapter 8 explores the practicability of image schemas as a design language in a human-centred design process. In study 7, image schemas are applied to all design phases: context-of-use analysis, requirement specification, producing design solutions, and evaluation. Furthermore, a database is described that aims at supporting designers in applying the image-schema language in different phases of a design process.

Chapter 9 wraps up the findings of the theoretical and empirical studies presented. The chapter reviews and evaluates the contribution image schemas make to designing intuitive interaction. It is discussed how image schemas relate to other guidance that has been suggested for supporting design for intuitive use. The outlook discusses requirements for further research.

Finally, the Appendix contains the detailed results of the data analyses.

2 Design for Intuitive Use

If the usefulness of image schemas as a design tool for intuitive use is to be assessed, it needs first to be clarified what intuitive use is and how it can be distinguished from the more widely established concept of usability (e.g. ISO, 1998). Against this background, the current guidance for designing intuitive use needs to be reviewed to determine its current benefits and limitations.

Although the term 'intuitive use' is widely used when describing interaction with products, only few attempts have been made to define the term in a scientific way. Generally, two approaches have been taken to arrive at a definition of intuitive use: (1) a top-down approach that reviews the literature about the phenomenon of intuition, and (2) a bottom-up approach that investigates the term 'intuitive use' as it is used by producers, users, and experts of human-technology interaction.

2.1 Top-Down View on Intuitive Use: The Literature

The best-known account of intuitive use is Jef Raskin's article "Intuitive Equals Familiar" (1994). Raskin states that intuitive use often means

> *that the interface works the way the user does, that normal human "intuition" suffices to use it, that neither training nor rational thought is necessary, and that it will be natural.* (p. 17)

'To intuit' means 'to suddenly understand with no apparent effort or reasoning'. Sometimes this makes 'intuitive' use a seemingly esoteric concept referring to supernatural abilities humans might or might not have. This might be the reason that researchers in human-technology interaction do not readily embrace the term 'intuitive use'. By looking at how naïve users use the mouse of an Apple Macintosh computer, Raskin comes to the conclusion that

> *a user interface feature is "intuitive" insofar as it resembles or is identical to something the user has already learned. In short, "intuitive" in this context is an almost exact synonym of "familiar".* (p.18)

The more familiar user interface elements are, the more readily they are understood and the less help is needed by the user. This view also allows for user interfaces becoming more intuitive or familiar with repeated use. Intuitive use, thus, is not something that can be measured only at the first interaction with a product. Raskin acknowledges that 'intuitive' means that users can readily transfer their existing skills to new user interfaces. But he also claims that user interfaces that are intuitively usable hamper progress. Focussing on such user interfaces means clinging to old and familiar interaction concepts that do not develop the full potential of technology for designing effective and efficient interaction.

Spool (2005) also refers to the users' knowledge in his account of intuitive use. He defines two conditions. In the first, users have exactly the amount of knowledge that is needed for solving the task with the system, i.e. the system is designed to match the knowledge level of the user population. In the second condition, there is a gap between the users' knowledge and the knowledge required to solve the task with the system. But the system will help the users to bridge the gap, although they are completely unaware of the fact. "The user is being trained but in a way that seems natural" (Spool, 2005).

While Spool (2005) and Raskin (1994) directly consider intuitive use from a perspective of human-technology interaction, intuition has been investigated in other disciplines as well. In their model of skill acquisition, Dreyfus, Dreyfus, and Athanasiou (1986) place intuition at the highest level of expertise. An expert, rather than relying on rules and explicit reasoning, will rely on tacit understanding or 'gut feelings' to judge situations. From a decision making perspective, Gigerenzer (2007) lists as characteristics of intuitive decisions that (1) they appear quickly in one's consciousness, (2) the decision maker is not fully aware of its underlying rationale, yet (3) the decisions feel strong enough to act upon.

The most comprehensive review of the literature on intuition as a background for defining 'intuitive interaction' was made by Blackler (2006). She reviewed the concept of intuition in creative thinking (drawing mainly on Bastick, 2003; Eysenck, 1995; Fischbein, 1987), decision making (Agor, 1986; Klein, 1993, 1998; Kolodner, 1993), memory and expertise (King & Clark, 2002; Schank, 1982), brain science (Damasio, 1994), consciousness studies (Baars, 1988), and education (Noddings & Shore, 1984). The conclusions of the review are that:

- Intuition is based on past experience "rather than on supernatural inspiration or some magical sixth sense" (Blackler, 2006, p.26).
- Similarly, in novel situations, people heavily rely on their past experience for decision making (cf. *case based reasoning*, Kolodner, 1993; and *recognition primed decisions*, Klein, 1993, 1998).
- Most of intuitive (vs. analytical) decision making is very quick and automatic, and often depends on simple heuristics.
- Prior experiences are stored in memory in form of scripts and schemas (Brewer, 1987) and are activated via spreading activation of similar neuronal connections (Hebb, 1949; Rumelhart & McClelland, 1986).
- Low conscious awareness is characteristic of intuitive processes, e.g. in implicit learning and decision making.
- A sudden insight, also known as the "Aha!" experience, occurs when subconscious thought leads to a build-up of information that at some point

becomes conscious (Bowden, 1997; Bowers, Regehr, Balthazard, & Parker, 1990) and that involves an affective response.

As a result, Blackler (2006, p. 74) defines intuition as

> *a type of cognitive processing that utilises knowledge gained through prior experience (stored experiential knowledge). It is a process that is often fast and is non-conscious, or at least not recallable or verbalisable.*

Consequently,

> *[i]ntuitive use of products involves utilising knowledge gained through other experience(s) (e.g. use of another product or something else). Intuitive interaction is fast and generally non-conscious, so that people would often be unable to explain how they made decisions during intuitive interaction* (Blackler, 2006, p. 120).

2.2 Bottom-Up View on Intuitive Use: How Producers, Users, and Designers Use the Term

The bottom-up approach on intuitive use has been taken by the IUUI research group (Intuitive Use of User Interfaces; Mohs, Hurtienne, Kindsmüller, Israel, Meyer, & the IUUI research group, 2006; Mohs, Hurtienne, Scholz, & Rötting, 2006). This group looked into the understanding of intuitive use amongst producers, users, and usability experts.

When analysing producer websites in which interactive products are advertised as *intuitive to use* the concepts most frequently associated with it are *usable without learning, training, or manual*; *simple and easy to use*; *efficient, fast, and direct to use (without detours)*; *consistent*; *usable without thinking*; *featuring (natural) mappings*; *clearly laid out*, and following the guideline *less is more*. By using the concept 'intuitive use' producers seem to emphasise minimal learning effort and simple and efficient use. Both, learnability and efficiency are also sub-concepts of usability as defined in the standards ISO 9241-110 and 9241-11. Producers and advertisers seem to use the concepts of 'usability' and 'intuitive use' almost interchangeably (Mohs, Hurtienne, Kindsmüller, et al., 2006).

When naïve users are asked what they understand by intuitive use, they say things like *action that is emotionally guided*, *to act on gut feelings*, or *to use without instruction or without explanation*. These and other expressions like *to use without thinking*, *automatic use*, or *routine use* point to intuitive use as an action that is not (any more) a cognitively demanding process. Intuitive use is done without external help and feels *logical*, i.e. conforms to user expectations (cf. the usability principle of 'conformity with user expectations' in ISO 9241-110). Intuitive use is seen as helpful during the first encounter of a product, but also helps using products the user is already familiar with (Mohs, Hurtienne, Scholz, et al., 2006). The focus of the users' understanding is narrower than the producers' understanding of intuitive use. Users talk more about the reduction of cognitive effort in operating a product; producers talk more about user interface surface characteristics (e.g. clear layout) or general efficiency.

Table 2.1. *Dialogue Principles according to ISO 9241-110 (ISO, 2006) and the Degree of Relatedness to the Concept of Intuitive Use*

Dialogue principle	Definition	Relatedness to Intuitive Use	
		M	*SD*
Conformity with user expectations	The dialogue corresponds to predictable contextual needs of the user and to commonly accepted conventions.	4.6	0.8
Self-descriptiveness	At any time it is obvious to the users which dialogue they are in, where they are within the dialogue, which actions can be taken and how they can be performed.	4.0	1.3
Suitability for the task	The dialogue supports the user in the effective and efficient completion of the task. The user is enabled to focus on the task itself rather than the technology chosen to perform that task.	3.5	1.2
Controllability	The user is able to initiate and control the direction and pace of the interaction until the point at which the goal has been met.	2.7	1.2
Suitability for learning	The dialogue supports and guides the user in learning to use the system.	2.4	1.2
Error tolerance	Despite evident errors in input, the intended result may be achieved with either no or minimal corrective action by the user. Error tolerance is achieved by means of damage control, error correction, or error management to cope with errors that occur.	2.2	1.1
Suitability for individualisation	Users can modify interaction and presentation of information to suit their individual capabilities and needs.	2.0	1.0

Note. Relatedness ratings were done by usability experts using five-point rating scales (1 = ‘not at all similar’ to 5 = ‘very similar’), *N* = 24.

Asking usability experts leads to almost the same results. User interfaces that are intuitive to use are seen as *conforming to user expectations* and *matching the users' mental model*. Also the expressions *easy to use*, *immediate* and *no thinking* were mentioned (Mohs, Hurtienne, Scholz, et al., 2006). At another occasion, usability and human factors experts were asked to relate ‘intuitive use’ to the seven dialogue principles of the ISO 9241-110 (ISO, 2006). Of the dialogue principles, ‘conformity with user expectations’, ‘self-descriptiveness’, and ‘suitability for the task’ were most closely associated with intuitive use. Medium to weak similarity judgements were made for the other dialogue principles: ‘controllability’, ‘suitability for learning’, ‘error tolerance’, and ‘suitability for individualisation’ (see Table 2.1 for mean values and standard deviations of expert’s ratings).

These analyses and a general review of usability criteria (Scholz, 2006) led to a tentative definition by the IUUI research group of intuitive use as the *users‘ subconscious application of prior knowledge that leads to effective interaction* (Mohs et al., 2006). Because of the vagueness of the term ‘subconscious’ it was later changed to a more operational definition: *intuitive use is given when users can apply their prior knowledge using a minimum of cognitive resources to effectively solving their task* (Hußlein et al., 2007).

Both definitions point out that there are preconditions and consequences of intuitive use. The preconditions are the use of prior knowledge and its subconscious processing. A consequence of intuitive use is the effectiveness of interaction. A second consequence follows from subconscious processing – intuitive use leads to using a minimum of cognitive resources. A third consequence, which is not explicitly stated, but which people always imply when talking about intuitive use, is the positive subjective evaluation of the interaction by the user (cf. Winkielman & Cacioppo, 2001).

The current definitions of intuitive use refer to the level of single operations that are either intuitive or not (Mohs et al., 2006). Often, however, people use the term ‘intuitive use’ when describing whole interaction episodes that aggregate across several operations. Hence, the definition should allow for a more gradual description of intuitive use going beyond simple yes/no statements.

Following these considerations, the following definition is proposed: *Intuitive use is defined as the extent to which a product can be used by subconsciously applying prior knowledge, resulting in an effective and satisfying interaction using a minimum of cognitive resources.* As intuitive use is an important and recurrent theme in this work, further explication of the preconditions and consequences specified in the definition is necessary. At first conscious versus subconscious processing is investigated in more detail, and then various sources of prior knowledge are discussed. As the defined consequences of intuitive use are very similar to the defined consequences of usability, these two concepts are compared thereafter.

2.3 Conscious and Subconscious Processes

The use of ‘subconscious’ refers to (1) low-level mental processes that are in principle inaccessible to consciousness and (2) high-level mental activities that are in principle available to consciousness but currently outside phenomenal awareness. The first class of processes mainly draws on procedural knowledge used to detect features of perceptual stimuli, to encode and decode language, and to perform routine motor tasks. The latter class mainly draws on declarative knowledge used in form of semantic knowledge about objects and episodic knowledge about events (Kihlstrom, 1984).

The distinction between conscious and subconscious processes reflects the distinction between controlled and automatic information processing. Controlled processing is bound to conscious awareness, while automatic processing is unavailable to consciousness. Controlled processing is limited in capacity and requires attention. It can be used flexibly in changing circumstances, but it is slow and sequential. Typical examples for a controlled process are deliberate problem solving, rational deductions, and reasoning with symbols.

Automatic processing is not limited in capacity and does not require attention. Automatic processing is fast and parallel. Automatic processes are very robust but inflexible. They always occur in a specific context upon a triggering stimulus, cannot be controlled, and are very hard to modify (Eysenck & Keane, 2005; Schneider & Chein, 2003; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). Typical examples for automatic processes are the early stages of perception, vegetative reactions, and large parts of motor control (e.g. maintaining balance). As the definition of intuitive use suggests, automatic and subconscious processes are an essential component of intuitive use.

2.3.1 A Model of a Human Information Processor

The distinction between conscious controlled and subconscious automatic information processing is reflected in the model of a *human information processor* by Rasmussen (1986). The model distinguishes between a conscious, symbolic, and sequential processor and a subconscious, analogue, and parallel processor. These processors are located above and beneath a 'threshold of consciousness' (the dashed line in *Figure 2.1*), respectively. A simplified version of the model is introduced here, and parts of it are further developed in the theoretical discussion about image schemas and their role in intuitive use.

Sensory input is automatically pre-processed in the *perception* unit of the model (*Figure 2.1*). The results are available to the sequential processor as well as to *goal control,* to the *dynamic world model,* and to the *mismatch detector*.

Goal control processes not only goals (in the conscious domain) but also emotions and biological needs (in the subconscious domain). Goal control can be influenced by the conscious processor as well as by automatic processes triggered by external stimuli (cf. Bargh & Ferguson, 2000).

The dynamic world model runs an internal simulation of the state of affairs in the external environment. It resides in long-term memory and mostly operates subconsciously. In human-technology interaction the environment, the user's body, and the interactive product are simulated in order to understand and to anticipate system interactions and responses. This enables users to actively make control decisions in real time. The dynamic world model is connected to *motor coordination* and thus allows for rapid feed-forward control of motor programs.

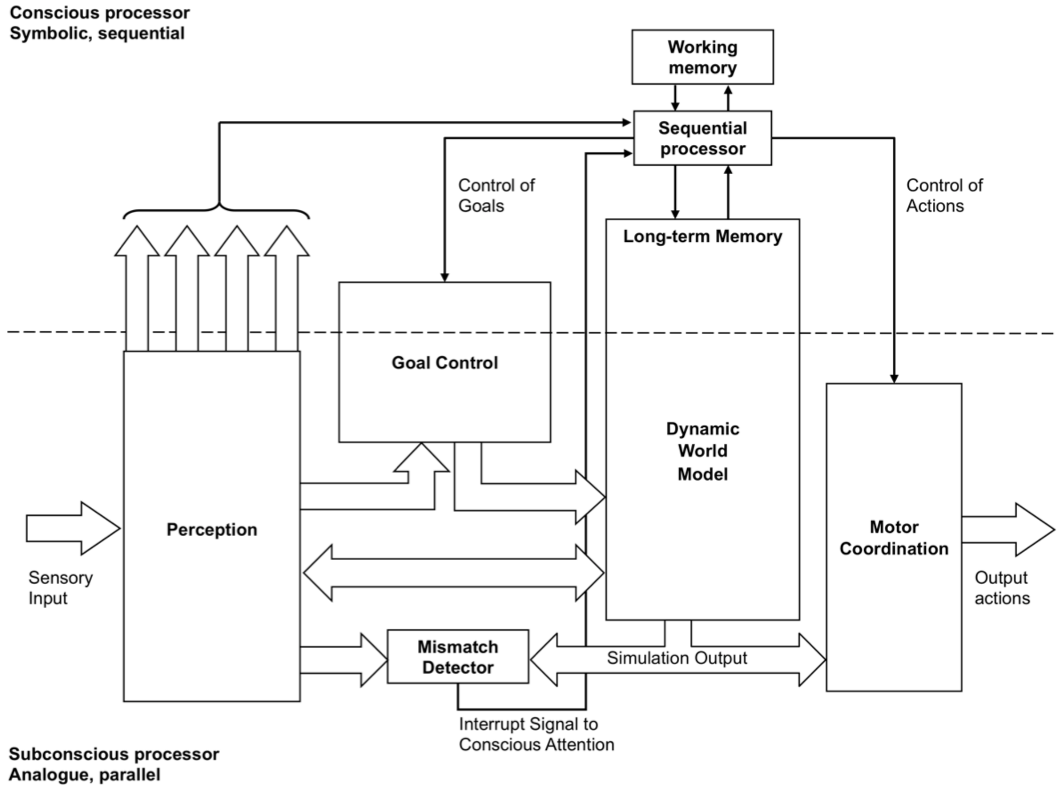


Figure 2.1. Model of a human information processor, after Rasmussen (1986, p. 76)

Whereas in the conscious domain the dynamic world model is something like a mental model, in the subconscious domain it is represented as automated rules and procedures. As discussed later, image-schematic representations are also part of the dynamic world model. Both imagination and complex problem solving are thought to involve the dynamic world model. The mismatch detector (*Figure 2.1*) alerts conscious control when the dynamic world model loses synchronism or is not properly updated due to unexpected behaviour in the environment.

Intuitive interaction involves the subconscious processor. Intuitive interaction is achieved when users do not have to reflect about the operation of the product (which would involve the conscious processor) and when they do not stop short because of mismatches between the flow of interaction with the product and the simulation of the interaction in the dynamic world model (because mismatches would trigger the mismatch detector that hands over to conscious control).

Published in 1986, the human information processor only needs minor updating to account for recent neuroscientific findings. These findings mainly imply that perception, motor coordination and the dynamic world model are more intertwined than the human information processor model suggests. An additional assumption, discussed in the following section, needs to be made to better understand the processing of image schemas and their metaphorical extensions within the model.

2.3.2 Partially Automatic Processing: Contention Scheduling and Schemas

With the idea of mismatch detection, Rasmussen draws a very distinct line between automatic and controlled processes. Other researchers have proposed a way of negotiating minor mismatches and resolving them within the subconscious processor. This seems to be a more useful approach in discussing the processing of image schemas than the fixed mismatch-based bisection in the Rasmussen model. Norman and Shallice (1986, cited after Eysenck & Keane, 2005, p.180) distinguish between three levels of functioning:

- fully automatic processing;
- partially automatic processing, involving a process called contention scheduling that works without deliberate direction or conscious control;
- deliberate control, by the supervisory attentional system (SAS), equivalent to the conscious processor.

While the first and the last level correspond to subconscious and conscious processing in the model of Rasmussen (1986), respectively, the second process, contention scheduling, is of interest here. It is assumed that when two routine activities are in conflict, contention scheduling will assess the relative importance of different actions and adjust the routine behaviour accordingly. For this, in memory, a hierarchical network of *schemas* is proposed that can also include image schemas on a lower representational level.

Schemas are abstract and represent objects or events stripped of their individual characteristics, retaining their common cores. Schemas can be embedded within each other. They are flexible and they represent knowledge at different levels of abstraction (Rumelhart & Ortony, 1977). On a low level, there are *motor response schemas* (like REACHING) that express relationships between a movement's initial conditions, response specifications, sensory consequences, and response outcomes (Cooper & Shallice, 2000). On a much more abstract level there are *scripts* (Schank & Abelson, 1977) that represent stereotypic sequences of events, e.g. a visit to a restaurant that involves the steps (1) arrival and seat-taking, (2) menu-reading and ordering, (3) eating, and (4) paying and leaving. In contrast to image schemas, scripts are encoded in amodal propositional form.

Image schemas are more abstract than motor response schemas, but far less detailed than scripts. They encode object and event characteristics that are informed by sensorimotor input, but on a higher level of abstraction than pure sensorimotor perception. Their encoding is still analogue and sensorimotor-like (cf. perceptual symbols, Barsalou, 1999).

Action regulation based on schemas is thought to function by spreading activation in the hierarchical network of schemas. Schemas are excited and inhibited by the actual goal state of the system, by the state of the environment, by the conscious processor (the SAS), and by the activity of other schemas (lateral or parental activation). When a schema exceeds a certain activity threshold and exceeds the activity of other schemas, it will get 'selected' and wins the contention scheduling process. This mechanism of partially automatic processing could take place within the dynamic world model of Rasmussen (1986), in which schemas can be representations arranged in a dynamic and hierarchical network.

The precondition of intuitive use was defined as *the subconscious application of prior knowledge*. To summarise, in this definition 'subconscious' means automatic or partially automatic processing via contention scheduling. Prior knowledge, including procedural knowledge and schemas, is seen as a part of the dynamic world model of Rasmussen's human operator. The same framework is applied in thinking about how image schemas function, when users interact with technology (chapter 3).

2.4 Acquisition of Knowledge

If intuitive use involves the subconscious application of prior knowledge, it must be considered how the application of prior knowledge becomes automated and what sources of knowledge there are.

Information processes that are automatic and subconscious can be acquired intentionally (as in learning to play an instrument) or unintentionally (as in learning to avoid touching a hot stove). Knowledge is applied automatically in a specific situation, if it was frequently and consistently used in similar situations in the past:

> *The necessary and sufficient ingredients for automation are frequency and consistency of use of the same set of component mental processes under the same circumstances – regardless of whether the frequency and consistency occur because of a desire to attain a skill, or whether they occur just because we have tended in the past to make the same choices or to do the same thing or to react emotionally or evaluatively in the same way each time.* (Bargh & Chartrand, 1999, p.469)

The process of automation becomes automatic itself when the learner is neither aware of the process nor of the results. According to Bargh and Chartrand (1999) this is how goals become tied to situations, and these situations might then automatically trigger the pursuit of these goals without any prior conscious selection. This could be the mechanism by which image schemas and their correlations with other concepts are automatically learned (e.g. associating *more* with UP and *less* with DOWN) and metaphorical extensions are formed (cf. chapter 3.1).

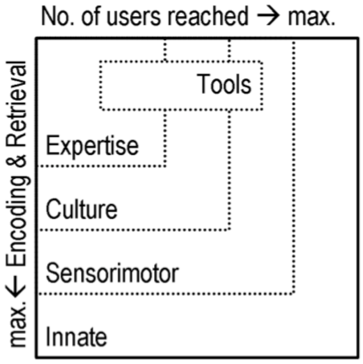


Figure 2.2. Continuum of knowledge sources in intuitive interaction

Whether automated intentionally or unintentionally, various sources of prior knowledge can be tapped. These knowledge sources can be roughly classified along a continuum (*Figure 2.2*).

The first, and lowest, level of the continuum consists of knowledge that is ‘acquired’ through the activation of genes or during the prenatal stage of development (*innate*). Generally, this is what reflexes and instinctive behaviour draw upon.

The next knowledge source is *sensorimotor*. Knowledge drawing on this source is to a large extent acquired early in childhood and is from then on used continuously through interaction with the world. Children learn for example to differentiate faces; they learn about gravitation; they build up concepts for speed and animation. Scientific concepts like affordances (Gibson, 1979) and image schemas (Johnson, 1987) refer to this level of knowledge.

The next level feeds into knowledge specific to the *culture* a person lives in. Underlying knowledge can vary considerably between cultures and may influence how people approach technology. This, for instance, touches the realm of values (e.g. what constitutes a taboo), the styles of visual communication (e.g. Japanese manga vs. American comics), but also knowledge about daily matters like the usual means of transportation (e.g. busses, trains, or bicycles) or the prevalent form of energy supply (e.g. by a public power line or by burning wood for heating).

The most specific knowledge source is *expertise*. It leads to specialist knowledge acquired in one’s profession, for example as a mechanical engineer, an air traffic controller, or a physician; as well as in hobbies (e.g. modelling, online-gaming, or serving as a fire-fighter).

Across the sensorimotor, culture, and expertise levels of knowledge sources, the knowledge acquired from using *tools* can be distinguished. Tool-based knowledge is an important reference when designing user interfaces. Tools at the sensorimotor level are primitive tools like sticks for extending one’s reach and stones used as weights. Tools at the culture level are those shared by many people, like ballpoint pens for writing or telephones for communication. At the last stage there is knowledge acquired from using tools in one's area of expertise, for example computer aided design (CAD) tools, enterprise resource planning (ERP) systems, or machine tools. Even within the same domain of expertise (e.g. engineering design) there may be differing knowledge on the tool level, depending on the kind of tools used (e.g. the software applications CATIA vs. Pro/ENGINEER).

The continuum of knowledge sources has an inherent dimensionality. The further towards the top level of the continuum, the higher the degree of specialisation of knowledge and the smaller the potential number of users possessing this knowledge. But still, on each level of the continuum 'intuitive use' (according to the above definition) can be assigned – as long as knowledge is subconsciously applied by users.

The application of knowledge may be subconscious from the beginning on (as with reflexes) or may have become subconscious due to very frequent exposure and reaction to stimuli in the environment. The more frequent the encoding and retrieval was in the past, the more likely it is that memorised knowledge is applied automatically. Both forms of knowledge acquisition apply, although unintentional acquisition might be more prevalent. Knowledge at the expertise level is acquired relatively late in life and is (over the life span) not as frequently used as knowledge from the culture or sensorimotor level. Expertise is likely to be acquired intentionally, although a large part of it may be acquired unintentionally and builds a deposit of so-called 'tacit knowledge'.

Knowledge from the lower levels of the continuum is more likely to be applied subconsciously than knowledge from the upper levels. Therefore, it is more likely to see intuitive interaction involving knowledge at the lower levels of the continuum. Limiting 'intuitive interaction' to the lower levels of the continuum has further advantages:

- The further down in the continuum, the larger and more heterogeneous are the user groups that can be reached. While almost everyone will have a concept of 'verticality' (sensorimotor level), not everyone understands the CATIA software package (tool/expertise level). This has another implication:
- Instead of being required to analyse the prior knowledge of the specific target user group, designers might simply refer to rules generated from findings about the general structure of human knowledge (i.e. general human knowledge on the sensorimotor level).
- Extremely frequent encoding and retrieval events lead to a higher robustness of information processing. In situations of high mental workload and stress a fall-back on lower stages of the knowledge continuum will occur. This will be especially important for the design of systems with high safety demands (e.g. control of an aircraft or of a nuclear power plant).

- Automatic processing of user interface elements in general goes along with lower workload on the conscious processor. Thus, more cognitive resources will be available for solving the (work) task at hand instead of wasting time and mental effort on figuring out how a piece of technology works.

Although the lower levels of the continuum of knowledge sources are emphasised by this approach, the definition of intuitive use still allows for intuitive use based on knowledge from higher levels that is well learned and can be applied automatically. Knowledge from the higher levels thus certainly helps expert users – but may come at the cost of robustness of interaction and the exclusion of non-expert users. This work will concentrate on image schemas at the sensorimotor level, but the reader should keep in mind that potentially all levels can be used in intuitive interaction.

After this discussion of the preconditions of intuitive use (subconscious processing and prior knowledge), the following subchapter considers the consequences of intuitive use and the options for measuring them. This requires comparing the concept of intuitive use with the concept of usability.

2.5 Intuitive Use as a Sub-Concept of Usability

It is time to clarify the distinction between usability as an established approach to designing human-technology interaction and intuitive use. Usability is defined as the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use (ISO 9241-11; ISO, 1998). Intuitive use is defined as the extent to which a product can be used by subconsciously applying prior knowledge, resulting in an effective and satisfying interaction using a minimum of cognitive resources (see above).

While the phrasings of the definitions of usability and intuitive use are different, the implications are similar. Intuitive use, like usability, is thought to take place in a specified context of use involving tasks, users, and technology (cf. Mohs et al., 2006; *Figure 2.3*). Intuitive use, like usability, is not a characteristic of the technology but of the degree of fit between the corners of the usability triangle in *Figure 2.3* (cf. Blackler, 2006; Spool, 2005).

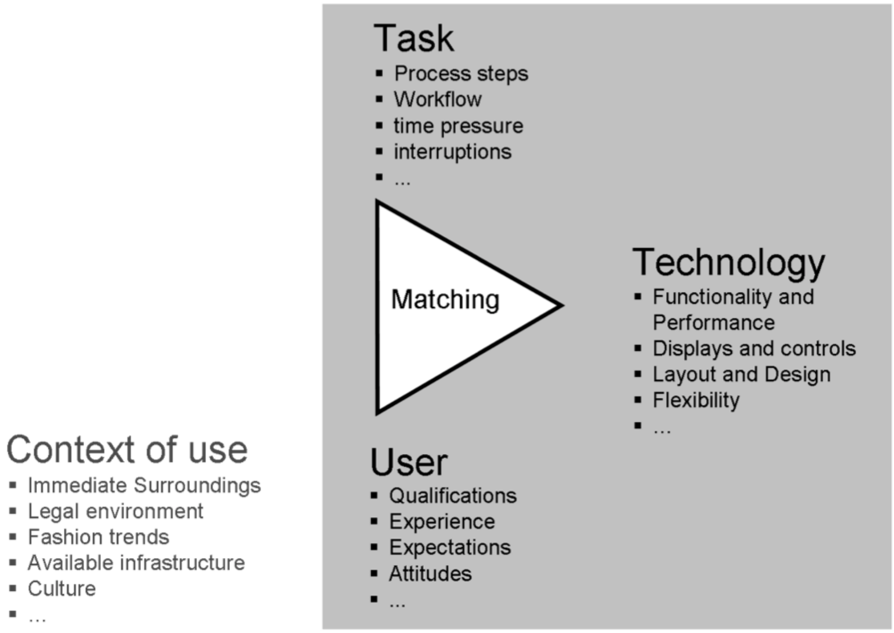


Figure 2.3. The Usability Triangle: Usability as the fit between users, their tasks, and the technology within a context of use (after Frese & Brodbeck, 1989)

Table 2.2 summarises the relevance of the three usability criteria effectiveness, efficiency, and satisfaction to design for intuitive use. The criterion of effectiveness is common to the definitions of both concepts. If users are not able to reach their goals with reasonable accuracy and completeness, the product is designed neither for usability nor for intuitive use. Effectiveness can be measured, e.g. by error rates or the quality of goal achievement.

The criterion of efficiency is also related to intuitive use. Intuitive use as the subconscious application of prior knowledge, however, only affects some of the indicators for efficiency. As the operational definition of intuitive use suggests, *mental effort* is the most notable of these, because intuitive use reduces the workload on the conscious cognitive processor. Mental effort can be measured either by objective or by subjective means. Objective measures include performance in dual task settings (e.g. the DisTracked method developed by Mohs, Lewandowitz, Patzlaff, & Jürgensohn, 2007) and physiological measures like heart rate variability (Rowe, Sibert, & Irwin, 1998). Subjective measures can be obtained from questionnaires that ask people how much effort they have spent achieving a goal, such as the Subjective Mental Effort Questionnaire (SMEQ, Zijlstra & van Doorn, 1985) or the NASA Task Load Index (TLX, Hart & Staveland, 1988). In the experiments described in chapters 5 and 6, a reaction time measure is employed that hints at changes in mental effort.

Table 2.2. *Usability Measures and their Relevance to Design for Intuitive Use*

Usability measures (ISO 9241-11)	Indicators	Relevance for Design for Intuitive Use
Effectiveness (the accuracy and completeness with which users achieve certain goals)	error rates	Yes (effective interaction is part of the definition of intuitive use)
	quality of goal achievement	
	proportion of users achieving goal	
Efficiency (the relation between the effectiveness and the resources expended in achieving the goals)	users' mental effort	Yes (indicators point to subconscious use of prior knowledge)
	number of references to help and documentation	
	learning time	
	users' physical effort	No (no direct correlation with intuitive use according to the above definition)
	task completion time	
	cost	
Satisfaction (the users' comfort with and positive attitudes towards the use of the system)	attitudes	Yes (users should be satisfied with using technology)
	preferences	
	subjective effectiveness and efficiency	
	experienced stress and strain	

The *number of references to help and documentation* is reciprocal to intuitive use, because it points out conscious processing. For first time interactions *learning time* may also be of interest. For further observation-based measures of intuitive interaction also see Blackler, Popovic, and Mahar (2007).

Efficiency measures that are not correlated with intuitive use are *task completion time*, *amount of physical effort,* as well as *cost*. 'Not correlated' means that the probabilities are equal for intuitive use to enhance, decrease, or not change the amount of these resources needed for using a specific user interface. Intuitive interaction is about subconscious processing and mental effort. Increased physical effort can in certain cases even support intuitive use, for instance by an increased number of mouse clicks.

Finally, the usability criterion of satisfaction is relevant, as it captures the experiential side of intuitive use. For measuring aspects of satisfaction that are specific to intuitive use, a standardised questionnaire is available (QUESI – questionnaire on the subjective consequences of intuitive use). The questionnaire comprises the subscales *subjective mental workload*, *perceived achievement of goals*, *perceived error rate*, *perceived effort of learning* and *familiarity* (Hurtienne, Dinsel, & Sturm, 2009; Naumann & Hurtienne, 2010).

The correspondence between usability and intuitive use shows that both concepts are closely related, but they are not the same. Intuitive use is a sub-concept of usability – only a subset of usability properties applies to intuitive use. Thus, design for intuitive use enhances the single aspect of mental effort at the possible cost of other efficiency measures. This specific focus of designing for intuitive use in comparison to designing for usability must be clear to user interface designers.

It is a commonly held belief that, while usability is relevant for virtually any interactive product, intuitive use is required especially for beginners, rare users, diverse user groups, or users that have no time or are unwilling to learn about operating a product. High frequency users, on the other hand, are said to want to achieve results quickly and to be happy with learning powerful functionality that is controlled by written command code, keyboard shortcuts, or transaction codes. According to this belief, intuitive use could stand in the way of the speedy interaction these users desire and may even be seen as a hindrance.

In this work, this view is not shared. The definition of intuitive use as the subconscious use of prior knowledge caters for novices and experts alike. While designing for novices needs to be based on knowledge at the lower levels of the knowledge continuum, designing for experts can also take into account automated knowledge from their area of expertise and the tools they use. If a keyboard shortcut is well learned and automated, it will be intuitive to use. The trends shaping the demand for intuitive use (cf. chapter 1), however, enhance the importance of designing for a novice and intermediate audience. Hence, the lower and more general knowledge sources are more suitable for addressing this demand and are the focus of this work.

2.6 Design for Intuitive Use: Current Guidance

Using our definition of intuitive use, the designer focuses on the search for design tools that address subconscious use of prior knowledge. Several tools and design criteria have been proposed for intuitive use (Blackler, 2006; Blackler & Hurtienne, 2007; Fischer, 1999; Mohs, Hurtienne, Scholz, et al., 2006). The most prominent among them are gestalt principles, affordance, compatibility, consistency, population stereotypes, and user interface metaphors (Table 2.3). These principles are not disjunct. They often overlap and can be nested within each other. Population stereotypes, for example, form a subclass of cognitive compatibility; affordance can be applied consistently or inconsistently across a user interface, etc.

Table 2.3. *Current Guidance for Design for Intuitive Use*

	Gestalt principles	Affordance	Compatibility	Consistency	Population stereotype	UI Metaphor
Description	Basic principles of perception that describe how a configuration or pattern of elements is seen as a unified whole.	Affordance is the quality of an object, or an environment that allows an individual to perform an action. *Physical* affordance depends on the capabilities of the actor. *Perceived* affordance also depends on the actors' experience, knowledge and culture.	Compatibility refers to corresponding locations, arrangements or movements of displays and their respective controls; or of UI elements to parts of the technical system. *Cognitive* compatibility exists when UI properties are congruent with user expectations acquired from system use or with general knowledge.	Consistency describes the uniformity of the appearance, location and/or behaviour of UI elements across different parts of the same system (*internal* consistency) or across different systems (*external* consistency).	Population stereotypes are mappings that are expected by most of the population. These mappings form response tendencies that are executed subconsciously and automatically. Population stereotypes either describe *spatial-to-spatial* mappings or *spatial-to-abstract* mappings.	UI metaphors allow the transference or mapping of knowledge from a source domain (familiar area of knowledge) to a target domain (unfamiliar area or situation), enabling users to use specific prior knowledge and experience for understanding and behaving in situations that are novel or unfamiliar.
Examples	Law of Proximity Law of Similarity	Physical affordance: the 'climbability' of stairs. Perceived affordance: the 'pushability' of a virtual button on a computer screen.	Moving a slider to the left lets the pointer in the display also move to the left (control-display compatibility).	Within a software application, the functions for copy and paste are always located at the same position in the menu (internal consistency).	Upward movement of a control for an upward or retract movement (spatial-to spatial). Clockwise movement for 'on' or 'increase' (spatial-to-abstract)	The Office/Desktop metaphor that transfers knowledge from using offices to the use of a computer operation system.

Table 2.3. *Current Guidance for Design for Intuitive Use (Continued)*

	Gestalt principles	Affordance	Compatibility	Consistency	Population stereotype	UI Metaphor
Benefits for designing intuitive use	Involve subconscious processing at the Perception stage Suggest concrete guidelines for UI design Possess a high degree of universality Contribute to a clear layout of the UI	Possess a high degree of universality Contribute to finding out 'what to do' with UI elements Are generally assumed to be processed subconsciously	Possess a high degree of universality Cognitive compatibility allows for physical-to-abstract mappings Contributes to UIs that look and behave in ways that seems logical to users.	Operations need to be learned and automated only once and then can be exploited in different situations Supports learning by making UIs more predictive Reduces the overall number of rules to be learnt	Allow for physical-to-abstract mappings *and* physical-to-physical mapping Contributes to operations that match users' expectations Give concrete design prescriptions	Users find familiar objects and behaviours in the UI Enhanced ease of learning and simplified use Motivation and stimulation Contribute to constraining the design space
Limitations for designing intuitive use	Scope is limited to basic arrangement and layout Only small effects on overall intuitive use are expected	More a philosophy than providing concrete design prescriptions Only perceived affordances can be used in software No support for handling abstract data	Different types of compatibility may be in conflict with one another Not much is known about how such conflicts can be resolved.	Consistency does not assure the proliferation of good design Consistency can hamper progress Consistent solutions can interfere with task requirements (e.g. to give priority to more frequent operations) Processing is not necessarily subconscious	Lack of theory about the origins of stereotypical mappings, thus no prediction of new mappings is possible Most stereotypes are based on asking people for the right mappings and performance studies are rare Only a limited number of stereotypes is known	Excludes users not familiar with the source domain of the metaphor Mismatches occur between source and target domain, i.e. unnecessary constraints or too many 'magical features' Mimicking known domains can hamper progress Not always subconscious processing

The current guidance on design for intuitive use can be evaluated along different lines. First it can be evaluated how well it fulfils the defined preconditions of intuitive use, i.e. whether it leads to subconscious processing based on prior knowledge. It can also be evaluated against the demand for intuitive use (cf. chapter 1). From an application viewpoint it is interesting to look at its scope specifying what and how much designers can achieve by applying it. Finally, it is interesting to look at the theoretical foundations of these principles, as the theory may provide hints on where to find new guidelines, and it enables the designer to explain the consequences of applying the principles in user interface designs.

Although the evaluation of these principles cannot be discussed in detail here, Table 2.3 hints at their benefits and limitations. Indeed, the principles differ in the degree to which they fulfil the defined preconditions of intuitive use. Gestalt principles, for example, are clearly processed automatically and subconsciously. At the other end of the spectrum there are user interface metaphors. These are not necessarily processed subconsciously and designers may even use them to enhance users' conscious learning of a user interface. All tools refer to some extent to prior knowledge, although on different levels. Gestalt principles refer to the innate level and affordances to the sensorimotor level. Compatibility guidelines mainly refer to the sensorimotor level, as well as population stereotypes. Consistency and user interface metaphors can draw on knowledge from each level of the continuum.

How well do the principles meet the demand for intuitive use? Some principles are near universal – as gestalt principles and affordance – and therefore can be used to accommodate heterogeneous user groups. Design principles like user interface metaphor and consistency that draw on different knowledge levels can also be used to cater for very specific user groups – but at the cost of general applicability if they refer to higher knowledge levels. Most principles can be applied to software and hardware alike, although physical affordance will only be found in hardware and more complex user interface metaphors will be mainly found in software. Also, the ability of providing means to cope with abstract data differs. While affordance is purely physical (or virtually physical in the case of perceived affordance), some population stereotypes and the principle of cognitive compatibility can also be applied to abstract data.

The design principles have different scopes. Gestalt principles, for example, contribute to a clear layout of the user interface (UI) and can determine which UI elements belong together and which do not. Affordances are limited to communicate how controls are operated. Consistency is a principle that only mimics other UI components but does not allow for much innovation. Of population stereotypes only a handful are documented in ergonomics textbooks. The widest scope and flexibility are found with user interface metaphors and the principle of compatibility. User interface metaphors can be used to simultaneously trigger knowledge about the appearance, location and behaviour of several UI elements. Cognitive compatibility, however, has the widest scope – although it is more a design philosophy than an approach giving concrete prescriptive guidelines for design.

Finally, the principles also differ in their quality and degree of theoretical foundation. Gestalt principles go back to early gestalt theory (e.g. Koffka, 1935). Affordance is a concept embedded in Gibson's *ecological approach to perception* (1979). Some general theoretical justification can be found for compatibility (Proctor & Vu, 2006) and (less) for consistency (Nielsen, 1989). Population stereotypes are mostly established as empirical facts without theoretical foundation. In texts about user interface metaphors conceptual metaphor theory (Lakoff & Johnson, 1980) is frequently mentioned (e.g. Neale & Carroll, 1997), but mostly without tapping the more concrete design derivations it can provide.

The discussion of the current design guidance for intuitive use shows that no single guiding concept exists so far that is appropriate in the widest range of usage situations. Besides its benefits, all known guidance has more or less serious limitations. Thus, it makes sense to look for further guidance, which might – at least to a certain extent – overcome these limitations. Image schemas have great promise to be such design guidance.

2.7 Summary and Conclusion

The first chapter described the demand for intuitive use and introduced image schemas as a promising design tool for intuitive use. This chapter defined intuitive use as the subconscious application of prior knowledge, resulting in an effective and satisfying interaction using a minimum of cognitive resources. According to this definition, subconscious processing is a basic precondition of intuitive use. Therefore, a model human information processor was discussed that differentiates conscious from subconscious processing. This model will provide the ground for locating image schemas within the information-processing stream as discussed in the next chapter.

As the application of prior knowledge is the second precondition of intuitive use, various sources of prior knowledge were differentiated and knowledge based on the sensorimotor level was identified as the most important factor concerning design for intuitive use. The next chapter argues the case for image schemas located at the sensorimotor level.

Regarding the consequences of intuitive use, intuitive use was identified as a sub-concept of usability. It was shown that intuitive use can be assessed via the usability criteria of effectiveness, mental efficiency, and satisfaction. If image schemas are posed to contribute to intuitive use, it needs to be shown that they can influence these criteria. The operationalisation of intuitive use will be relevant in the experimental and empirical studies described in chapters 5, 6 and 8 as well as in judging previous applications of image schemas in user interface design (chapter 4).

Finally, the brief review of the current guidance on designing for intuitive use showed the criteria against which any new design principle has to be evaluated. The new guidance needs to fulfil the defined preconditions for intuitive use as well as the market demand for intuitive use. Its scope should be as wide as possible and it should rest on a coherent theoretical foundation. The current design guidance fulfils these criteria to different degrees. In the next chapter the current evidence on how well image schemas can contribute to these criteria is reviewed. Their comparison to the current guidance for achieving intuitive use is taken up in chapter 9 as part of reviewing the empirical evidence created in this work.

3 Image Schemas and Intuitive Use

The same criteria used to judge the current guidance for designing intuitive interaction need to be applied to judging image schemas as design guidance. This means to address the theoretical foundation of image schemas, to consider how well image schemas meet the preconditions of intuitive use and to review how well they meet the demand for intuitive use.

3.1 Theoretical Foundation

Image schemas are a concept that is new to usability research – unlike the principles for designing intuitive use discussed in the last chapter. The origins of image-schema theory lie in philosophy and cognitive linguistics and are deeply connected to the notion of conceptual metaphor. Both concepts, image schemas and conceptual metaphors, are discussed in sequence.

3.1.1 Image Schemas

The term *image schema* was introduced by the philosopher Mark Johnson: "An image schema is a recurring, dynamic pattern of perceptual interactions and motor programs that gives coherence and structure to our experience" (1987, p. xiv). Applying phenomenological analysis to basic sensorimotor everyday experiences, Johnson derives 29 image schemas. One example is the image schema of a CONTAINER. Consider the many experiences of containers in daily activities like these:

> *Take for example a child in a red dress who watches her mother put cookies into a jar. The child then takes the lid off the jar and looks inside to search for the cookies. She reaches into the jar, reaches down into the cookies to find a particular cookie near the bottom, grasps the cookie (so that the cookie is now in her hand), and takes it out. She wraps the cookie in a napkin. She walks with the cookie through a door into another room, where she is picked up in her mother's arms and put into a high chair. She watches the mother pour milk into a glass. She then dunks her cookie into the milk (which is itself contained in the glass), and puts the cookie into her mouth*. (Dewell, 2005, p. 371-372)

Recurring and similar interactions with the world leave traces of these experiences in the brain (*Figure 3.1*). Crucially, these traces bear a resemblance to the perceptual and action processes that generated them, and are highly abstract. The experience with containers of all kinds, for example jars, glasses, clothes, rooms, and our own bodies, results in the mental representation of a CONTAINER image schema. Each CONTAINER can be described formally as consisting of an inside and an outside, separated by a boundary.

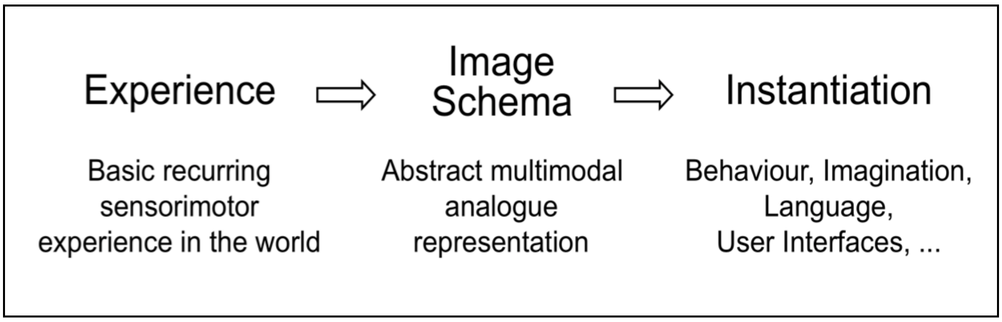


Figure 3.1. Acquisition, representation, and instantiation of image schemas

Once an image schema is formed, it can be instantiated in different ways. The CONTAINER image schema, for example, can be instantiated by a user interface designer when designing a field for text entry or when naming a part of an email application *inbox*.

According to Johnson (1987), image schemas are more than a simple decomposition of their parts implies. Image schemas are holistic gestalts (Lakoff, 1987; Johnson, 1987). Image schemas obtain their holistic character from the experiential entailments that form part of their meaning. The entailments of the CONTAINER image schema, for example, are: (1) A container provides protection from or resistance to external forces, e.g. the cookies in the closed jar are protected against humidity and other environmental influences. (2) Forces within the container are limited and restricted, e.g. the milk will not flow out of the glass. (3) The content's location is determined by the location of the container, e.g. the cookies move with the jar when the jar is moved. (4) The container makes its content more or less accessible to view (depending on the opening and the material). (5) Containment is transitive: if the cookie is in the napkin and the napkin is in the hand of the child, then the cookie is in the child's hand (Johnson, 1987).

The experiential acquisition of other image schemas is similar: The foundation of the UP-DOWN image schema is the experience of gravity. The image schema NEAR-FAR derives from the experience of reaching (cf. chapter 6), and PATH develops from experiencing trajectories of movement with their start and end points. Other image schemas derive from force-dynamic experiences like ATTRACTION, BLOCKAGE, DIVERSION, or are derived from prevalent object characteristics like BIG-SMALL, BRIGHT-DARK, HEAVY-LIGHT. For the current set of image schemas used in this work see chapter 4 (Table 4.1).

Image schemas share some basic properties:

- Image schemas are analogue representations deriving from experience. Analogue means that image schemas mirror the sensory experience, although they are much more abstract. In this respect they are extreme summaries of perceptual states – in contrast to abstract symbolic representations.
- Most image schemas have a basic structure. The CONTAINER image schema consists of an inside, an outside, and a boundary. The PATH image schema consists of a starting point, a trajectory, and an end point. Different parts of the structure of an image schema can be focussed and the focus can shift. In the image schema PATH the focus can be on the starting point, the end point, or the trajectory. This basic structure of image schemas is important for metaphorical mappings (see below).
- Image schemas have static and dynamic aspects. For instance, the NEAR-FAR image schema can denote static locations, i.e. whether an object is NEAR or FAR from a reference object, a person's body, or the current centre of attention. In its dynamic version it encodes movement, i.e. movement towards (from FAR to NEAR) or away (from NEAR to FAR).
- Image schemas are multimodal. Image schemas are not restricted to the visual modality. They also involve the auditory modality, as well as the touch and kinaesthetic senses.
- Image schemas operate beneath conscious awareness. In terms of Rasmussen's human information processor (chapter 2) image schemas reside in the dynamic world model and are used for simulating the state and dynamics of the world, guide perception, and prepare actions (cf. below, 3.2).

Many image schemas Johnson (1987) derives from his phenomenological analyses show great overlap with the findings of cognitive linguists who investigated the cognitive foundations of grammar and semantics. Major linguistic findings include (1) that a finite set of spatial categories, resembling Johnson's SPACE image schemas, repeatedly occurs in the meaning of prepositions across different languages (Levinson & Meira, 2003; Talmy, 1983, 2005); (2) that the description of force-dynamic events in language follows very simple patterns resembling Johnson's FORCE image schemas (Talmy, 1988); and (3) that in many African languages that only have non-extendable sets of between 2 and 27 adjectives, a set of core adjectives is common to these languages, which are comparable to the group of ATTRIBUTE image schemas (Dixon, 1982; Segerer, 2008). These analyses led to the conclusion that these repeated patterns in and across languages describing basic properties of objects and their relations are more than random concurrences. These patterns rather point to underlying cognitive structures – a foundational assumption of cognitive linguistics.

Apart from language, image schemas are thought to be involved in structuring perception, imagination, and motor action. Image schemas can also be instantiated in physical artefacts and their user interfaces. In a form filling dialogue, for example, each text box instantiates a CONTAINER image schema. The travel routes displayed in Google Maps instantiate the PATH image schema. The landing gear lever in an Airbus A320 instantiates the UP-DOWN image schema: when the pilot moves the lever downward, the undercarriage extends downwards; when the pilot moves it upward, the undercarriage retracts upward into the plane's body. Thus, one application of image schemas in user interfaces is for simple physical-to-physical mappings. To the designer, however, more interesting than physical-to-physical mappings are physical-to-abstract mappings. For these kinds of mappings, the idea of *conceptual metaphor* needs to be discussed.

3.1.2 Conceptual Metaphors

As mentioned earlier, image schemas can be extended to abstract domains. These metaphorical extensions are a subgroup of what has been called 'conceptual metaphors'. The notion of conceptual metaphor has first been developed in cognitive linguistics research.

In traditional linguistics, metaphor is defined as a figure of speech in which an expression is used to refer to something that it does not literally denote, as in *my love is a fragile flower.* According to the traditional approach, metaphor functions as rhetoric decoration or embellishment that, although not meant literally, can be easily paraphrased into literal language. Conceptual metaphor theory takes a different stand. In contrast to the traditional approach that focuses on novel or poetic metaphor, it is primarily concerned with so-called 'dead metaphors', i.e. metaphors that are so common that they normally go unnoticed in everyday parlance.

In examining a broad range of linguistic expressions, Lakoff and Johnson (1980) found that metaphor is a very frequent and regular feature of language. Moreover, they found that different metaphorical expressions can be clustered together forming systematic mappings from one domain to another. Consider the expressions *Things are looking up / This is a high-quality product / We hit a peak last year, but it's been downhill ever since.* All of these occur in standard language and their meaning cannot be understood literally. In all expressions there seems to be a mapping from the vertical dimension to the domain of quality in a form that can be described as GOOD IS UP – BAD IS DOWN. Lakoff and Johnson (1980) conclude that such clusters of metaphorical expressions are far too common and systematic that they could just be devices of language. Instead, these metaphors must be devices of the underlying conceptual system in the mind.

While the example GOOD IS UP – BAD IS DOWN involves a mapping of the image schema UP-DOWN to the more abstract domain of quality, Lakoff and Johnson (1980) were concerned with all kinds of metaphors, not just image-schematic ones. An example of a more complex metaphor is LOVE IS A JOURNEY that instantiates in expressions like *Look how far we've come. I don't think this relationship is going anywhere. It's a dead-end street.*

In the vast majority of metaphors the mapping between domains is unidirectional. Typically, mappings occur from a concrete source domain to an abstract target domain, from the physical to the non-physical, the more clearly delineated to the less clearly delineated (Lakoff & Johnson, 1980). Indeed, it proves rather difficult to reverse the direction of these mappings, e.g. talking about journeys in terms of love. Empirical evidence shows that concrete-to-abstract mappings are rated as more comprehensible and apt than abstract-to-concrete mappings and abstract-to-abstract mappings. Moreover, listeners highly agree in their paraphrases of metaphors with concrete-to-abstract mappings, while paraphrases of other metaphors are widely heterogeneous (Jäkel, 2003). More evidence for the unidirectionality hypothesis comes from etymologic studies that find that polysemous words like *heavy* first start out with concrete meanings (*heavy* referring to weight) and later acquire abstract meanings (*heavy* denoting difficulty as in *things were getting pretty heavy*) but not vice versa. For other examples see Haser (2003), Jäkel (2003), Sweetser (1990), Traugott and Dasher (2002).

Image schemas play a double role in the acquisition and structuring of metaphors. First, they are the source domain of many *primary metaphors*. Second, image schemas form the basic structures of more complex metaphorical mappings.

Image Schemas in Primary Metaphors

Primary metaphors arise from repeated correlations of concrete physical sensorimotor experiences and more abstract subjective experiences or judgements (Grady, 1997a, 1997b). The theory of primary metaphor can thus explain the association of image schemas with abstract concepts (*Figure 3.2*). The vertical level of a liquid in a container, for example, correlates with the quantity of the liquid; the amount of paper in a pile correlates with the vertical extension of the pile, and so on. Hence, in many contexts, quantity is connected to verticality. In learning about the world as children, these connections between domains are automatically learned as well. Through repeated experience with different quantities in different contexts, these connections become generalised. As a result, verticality is connected with quantities of all sorts – including non-physical quantities that are conceptualised on an UP-DOWN axis like the linguistic metaphors *The inflation is rising* or *The gross domestic product is at an all time low.*

In this example it seems that the image schema UP-DOWN is recruited for the conceptualisation of the more abstract concept of quantity. This conceptual connection between the domains of verticality and quantity is the conceptual metaphor, sometimes also called the metaphorical extension of an image schema. The notation of conceptual metaphors follows the convention TARGET DOMAIN IS SOURCE DOMAIN, hence in the example: MORE IS UP – LESS IS DOWN.

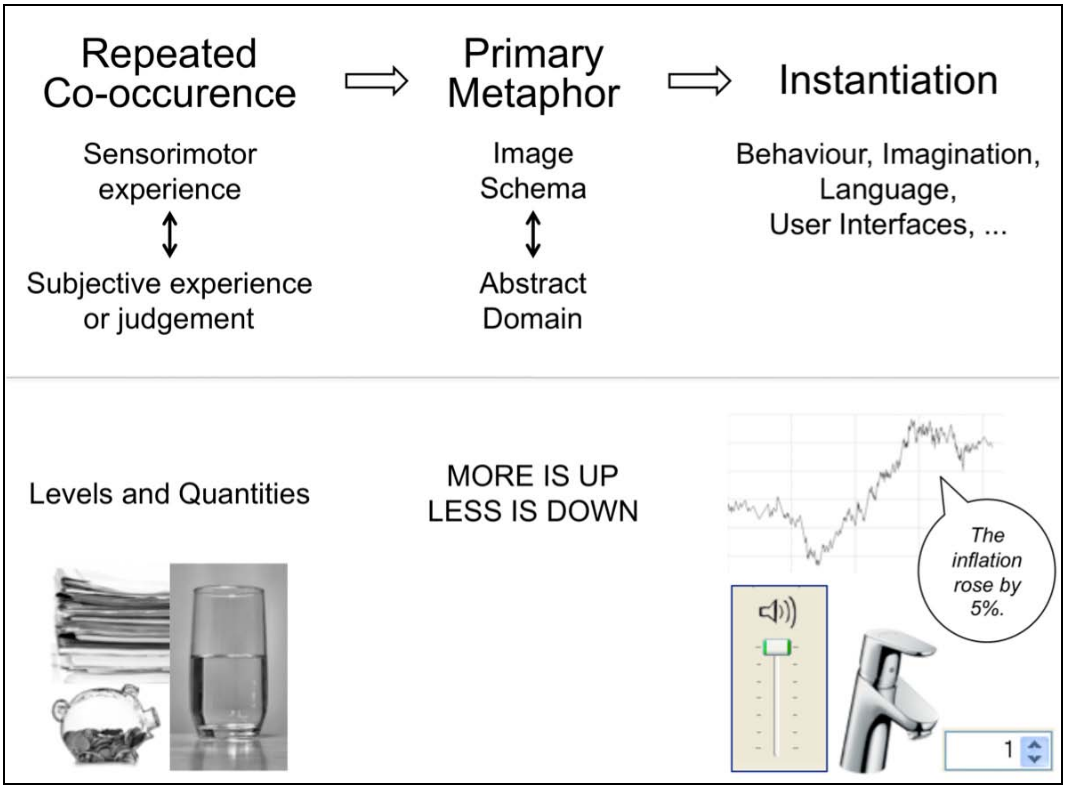


Figure 3.2. Acquisition and instantiation of primary conceptual metaphors, with examples

Like image schemas, conceptual metaphors are assumed to operate subconsciously and are instantiated in behaviour, imagination, language, and eventually user interfaces (*Figure 3.2*). The conceptual metaphor MORE IS UP – LESS IS DOWN, for example, is instantiated in metaphorical linguistic expressions like *the inflation rose by 5%* or *He is underage*. It also is instantiated visually in charts, for example showing the development of share prices of a company. In user interfaces MORE IS UP – LESS IS DOWN can be found in a vertical slider controlling the volume of the speakers, a water tap, or a spin box.

Other correlations in experience form other image-schematic conceptual metaphors. For instance, when carrying heavy objects, the sensory judgment of an object's mass is correlated with affective states associated with exertion. With repeated experience, the domains of exertion/difficulty and heaviness become mentally connected forming the conceptual metaphor DIFFICULT IS HEAVY. The metaphor is instantiated, for example, when moaning about the *burden of work to do*. Similarly, intimacy co-occurs with physical closeness forming the metaphor INTIMACY IS CLOSENESS drawing on the NEAR-FAR image schema. It is instantiated in sentences like *I'm very close to him*.

Image Schemas in Complex Metaphors

Primary metaphors are acquired via correlations in sensorimotor experience, and most primary metaphors are image-schematic metaphors. In user interface design, they can be useful for simple physical-to-abstract mappings. But user interfaces tend to be more complex than that. More complex metaphors, and also novel metaphors, do not necessarily rely on correlations in sensorimotor experience, but may nevertheless recruit image schemas for their internal structuring. In cognitive linguistics this idea is expressed in the *invariance hypothesis*.

The invariance hypothesis states that in metaphorical mappings the image-schematic structure of the source domain is preserved. This hypothesis is based on the observation that in complex metaphors only some aspects of the source domain are mapped to the target domain and other mappings are left out. The result is that some aspects of the target domain are highlighted and others are hidden (Lakoff & Johnson, 1980). When the metaphor LOVE IS A JOURNEY is used for describing a love relationship, it highlights the course of time that is shared by the lovers. Another metaphor, LOVE IS UNITY instantiated in *We were made for each other. She is my better half. Theirs is a perfect match*, highlights the mutual dependency of the lovers, but abstracts away from the time-course of the relationship (Lakoff & Johnson, 1980).

What determines the highlighting and hiding of elements of the target domain? The invariance hypothesis provides an answer: “Metaphorical mappings preserve the cognitive topology (this is, the image-schema structure) of the source domain” and “a great many, if not all, abstract inferences are actually metaphorical versions of spatial inferences that are inherent in the topological structure of image schemas” (Lakoff, 1990, p. 54). This means, for example, that the LOVE IS A JOURNEY metaphor preserves the structure of the PATH image schema highlighting the common trajectory and goal of the lovers. Likewise, the LOVE IS UNITY metaphor preserves the structure of the PART-WHOLE image schema and highlights the connectedness of separate individuals – the WHOLE forming from the two PARTS.

The invariance hypothesis implies that not all kinds of metaphors are acquired via a process of repeated occurrences in experience as suggested in *Figure 3.2*. Especially novel metaphorical expressions like *my job is a jail* cannot be explained this way if no experience of being in a jail exists. Yet it is easy to find the meaning of the sentence: the job is unpleasant and confining. Note, that other interpretations are possible, but are usually not invoked, e.g. that the food in the cafeteria is served in tin bowls or that there are steel bars in front of the windows. The interpretation made draws on the CONTAINER image schema and its entailments. One of the entailments is that the content of a container is restricted in movement. If the content is a person that is restricted in literal or metaphorical movement, this situation is perceived as unpleasant and confining.

The invariance hypothesis has two important consequences for user interface design. First, by providing the structure of metaphorical mappings it can help to focus the designer on the relevant features in user interface metaphors – image schemas determine what is highlighted and what is hidden. This is shown in *Figure 3.3*. Here, in the user interface the metaphor of a trashcan instantiates the mapping DELETING FILES IS USING A TRASHCAN. At the heart of the mapping is the structure conveyed by the image schemas PATH and COMPULSION (throwing / dragging files into the trash can) as well as CONTAINER and FULL-EMPTY (the trashcan as a CONTAINER holds files, shows different states of being FULL or EMPTY, and files can be retrieved if needed). However, the concrete instantiation of the image schemas, e.g. whether the CONTAINER takes the form of a paper basket or a garbage bin, can be determined by the designer based on the specific circumstances of the usage situation. The metaphor and image schemas should then also be present in the interaction with the interface (e.g. implementing the PATH image schema as drag-and-drop) and should be able to be detected in the user's language (e.g. *put the file into the trash*).

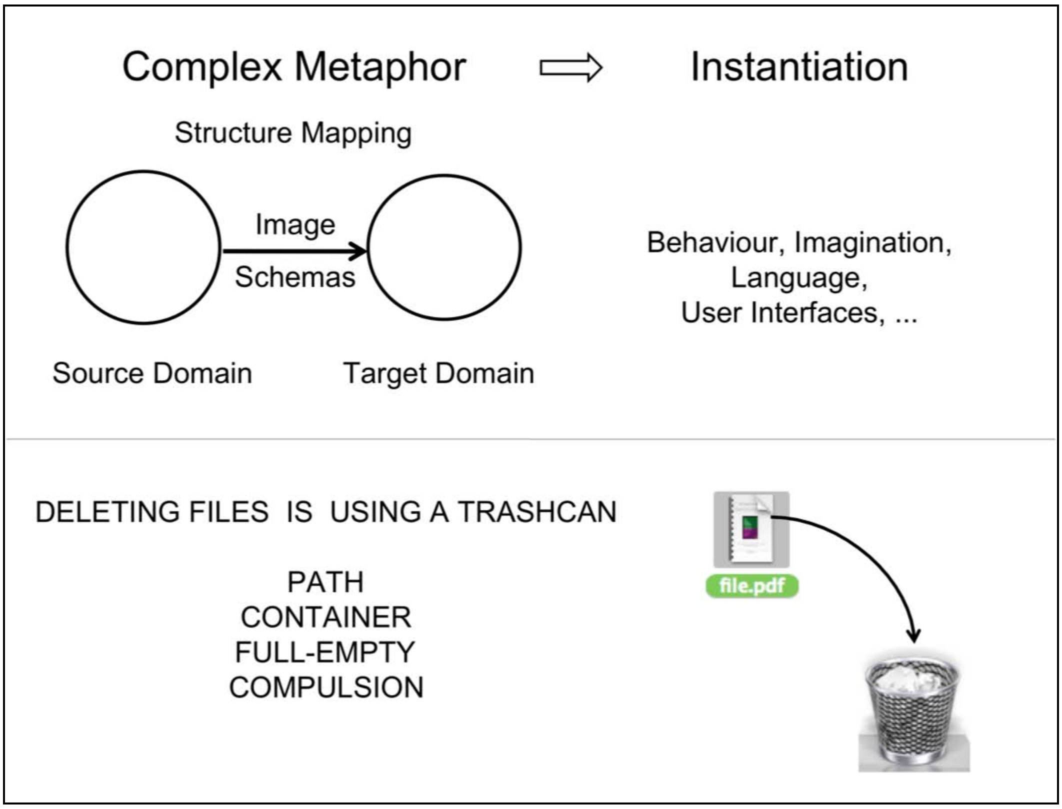


Figure 3.3. Image Schemas Providing the Structure of Complex Metaphors

Second, if thinking about abstract domains is hardly possible without using conceptual metaphors (Lakoff & Johnson, 1980, 1999), and if image schemas structure most metaphorical mappings (the invariance hypothesis), then image schemas should be regularly involved in thinking about abstract domains. This is an addition to the claim that image schemas describe basic patterns of physical sensorimotor experience (see above).

This ubiquity of image-schemas in thought and, consequently, in language, behaviour, and user interfaces can be exploited in designing an image-schematic metalanguage for user interface design. Then, image schemas can be used to describe user's mental models, their tasks, and user interfaces, independent from pre-configured metaphorical analysis. The idea of image schemas as a metalanguage follows from the theories reviewed in this chapter and from applying the basic principle of intuitive use from chapter 1 – that user interfaces should match the mental models of users. Whether it is feasible and useful to employ image schemas as a metalanguage for designing user interfaces is explored in the empirical part of this work that apply image schemas in a human-centred design process.

Concerns about the Psychological Reality of Image Schemas and their Metaphorical Extensions

The main controversy about image schemas and their metaphorical extensions concerns their psychological reality. Linguists claim to have found *conceptual* metaphors but only provide linguistic expressions as support for their claim. Inevitably this leads to circular reasoning (Glucksberg & McGlone, 2001). Cognitive linguists, for example, would argue that people think about love relationships in terms of journeys, because people often talk about love relationships using journey-related expressions. But then, to explain their findings they state that people often talk about love relationships using journey-related expressions, because people think about love relationships in terms of journeys.

Despite the great systematicity of metaphorical expressions in language it could be that the conceptual metaphors extracted by cognitive linguists are mere linguistic phenomena. Stronger evidence is needed that suggests non-linguistic forms of metaphorical processing (Peeters, 2001) to confirm the claims of the conceptual nature of metaphors. Only then the promise of conceptual metaphor for user interface design can be substantiated. Although several studies confirm the psychological reality of conceptual metaphor involving non-linguistic stimuli, a broad range of conceptual metaphors still awaits empirical investigation (cf. section 3.3.1).

Another issue that puts cognitive psychologists at unease concerns the internal validity of present linguistic studies (Glucksberg & McGlone, 2001; Murphy, 1996; Valenzuela & Soriano, 2005). The concern is about the methodology used for deriving metaphorical mappings from linguistic data. Often, as in the studies of Lakoff & Johnson (1980), it is not clear where linguistic expressions originated from. It would be important, however, to know the communicative contexts in which these expressions were used. In journalistic texts, metaphors might be more consciously used as an instrument of rhetoric than in everyday speech where mappings are more likely to be subconscious. Further, it is not exactly clear how

Lakoff and Johnson (1980) arrive at specific conceptual metaphors. This issue is important for understanding the method applied and acknowledging the validity of the results.

Both issues have been addressed in the linguistic community. Most metaphor analyses today are based on pre-defined text corpora; and the method of extracting metaphors is made explicit (cf. Baldauf, 1997; Goschler, 2008; Jäkel, 2003; Steen, 1999; Stefanowitsch, 2006). Further, quantitative statistical approaches begin to complement the usual qualitative approaches to metaphor analysis (cf. Stefanowitsch, 2006). However, the core of most studies, the extraction of conceptual metaphors, remains subjective, because authors mainly rely on their own intuitions in this process. In most cases only one analyst per study is involved, although converging evidence from more than one analyst would enhance the reliability of the metaphors extracted. Although evidence converges across different authors in the literature, it cannot be excluded that analysts are familiar with metaphors found in other studies and that they just replicate these metaphors in their own findings. Taken together, these problems still emphasise the need for a more objective and independent study of metaphor. They also put a grain of caution into an otherwise promising approach.

Other controversies surrounding conceptual metaphor theory include the specific format of mental representation of conceptual metaphors (cf. Murphy, 1996) and the proclaimed universality of conceptual metaphor across languages and cultures (see section 3.3.2).

3.1.3 Summary and Conclusion

The previous discussion has shown that image schemas are well grounded in theory. Image schemas are abstract and analogue representations of recurring physical sensorimotor experience. Once formed, these representations can be instantiated in language, in behaviour, and in user interfaces. Although derived from physical experiences, image schemas are also recruited for understanding more abstract domains. They are involved in primary metaphors that form from repeated correlations of otherwise distinct experiences and that are subsequently generalised from concrete to abstract domains. Image schemas are also involved in more complex metaphors. According to the invariance hypothesis, image schemas provide the structure that is mapped from the source to the target domain of a complex metaphor.

The theory implies that instantiations of image schemas can be applied to physical-to-physical and physical-to-abstract mappings in user interfaces. In particular, the invariance hypothesis allows for a more general application of image schemas as a structural description language in user interface design.

While image-schema theory looks very promising in its ramifications for user interface design, most evidence comes from linguistic studies. Although the theory makes a plausible case for the conceptual status of image schemas and their metaphorical extensions, the original accounts do not offer much empirical evidence for it. Thus, in the following, evidence is discussed on how image schemas are able to fulfil the preconditions of intuitive use and the promises for user interface design.

3.2 Image Schemas and the Preconditions for Intuitive Use

To deliver successful guidance for design for intuitive use, image schemas need to meet the preconditions for intuitive use. The precondition for intuitive use is derived from the definition of intuitive use as the subconscious application of prior knowledge. Prior knowledge can derive from different sources that can be innate, sensorimotor, culture, and expertise (chapter 2.4). Its subconscious processing means that prior knowledge is automatically activated and, although not attended to, becomes relevant for action. Image-schema theory claims that image schemas are forms of prior knowledge at the sensorimotor level and that they are processed subconsciously and therefore should be activated automatically. Are these hypotheses supported by empirical evidence?

3.2.1 Prior Knowledge

If image schemas are a form of sensorimotor knowledge representation, there needs to be an account of how image schemas derive from sensorimotor experience and of how they are connected to other forms of representation used in language and conscious thought.

According to Mandler (1992, 2004, 2005), image-schematic representations are already forming during the first year of life – long before children start to use language for communication. Image schemas are the result of a process called perceptual meaning analysis (PMA). In perceptual meaning analysis, perceptual information is attentively analysed and conceptual information (i.e. meaning) is abstracted from it. This process of PMA is necessary to reduce the vast amount of information processed in the parallel processor to an amount of information that can be handled by the conscious processor with its limited resources (Mandler, 2005). The representational format of the abstracted perceptual information is the image schema. Although not accessible to consciousness, image schemas then are involved in structuring the concepts accessible to consciousness, either in the form of images or words.

Perceptual meaning analysis can already be detected in infants who are three months old (Mandler, 1992; Werner & Kaplan, 1963). PMA in the infant is guided by an inherent bias towards moving stimuli and a yet underdeveloped sensorimotor system. Thus, the first things newborns see are blurry objects moving along trajectories. The PATH of such motion may be the first image schema that arises, and with further PMA more detail will be added. PATHs start at some point in space and have an endpoint where the motion comes to rest. The

PATH information is extracted from the perceptual display without representing information about speed, direction of movement, or any specific details about the objects involved.

Further perceptual meaning analysis reveals that there are different possibilities of how motion trajectories start. Entities can move by themselves or can be caused to move by other moving entities, thus constituting SELF-MOTION and COMPULSION image schemas. The distinction between these two image schemas can be shown empirically in infants that are six-months old (Leslie, 1982). With further PMA of motion events the infant learns that objects move along PATHs of different forms (STRAIGHT or NON-STRAIGHT), or that movement can be dependent, or contingent, on his or her own behaviour (contingency leading to the image schema LINK, Mandler, 1992, 2004). Further, the entailments of image schemas are learnt gradually by the infant with continued PMA. Fore example, the different entailments of the CONTAINER image schema, e.g. the co-location of the content and container, develop between 2.5 and 7.5 months of age (Baillargeon & Wang, 2002; Hespos & Baillargeon, 2001).

Further studies suggest that image schemas are the basis on which infants build more complex pre-verbal concepts. The concept of *support*, for instance, is built on the image schema CONTACT, available to three months old infants, to which later, at five months the image schema UP-DOWN, and at six and a half months the image schema BALANCE is added (Baillargeon, Kotovsky, & Needham, 1995).

Image schemas and these early concepts also provide the basis for the acquisition of language. Image schemas in particular ease the learning of terms denoting spatial relations that are more difficult to point out than objects. Across different languages, for example, the first spatial preposition children learn is *in* (Johnston, 1988). Image schemas like CONTAINER and IN-OUT are already there as meaningful partitions: “What remains for children to do is to discover how their language expresses these partitions” (Mandler, 1992, p. 599). Further, locative terms like *on*, *under*, *next to / beside, back* and *front* are acquired in a predefined sequence (Johnston, 1988) based on the image-schematic concepts: ‘support’ (a combination of CONTACT, UP-DOWN, BALANCE), UP-DOWN, NEAR-FAR, and FRONT-BACK, respectively. From the errors children typically make in the acquisition of their respective mother tongues it can be derived that the use of abstract grammatical markings is also based on image schemas (Choi & Bowerman, 1991; Mandler, 2004).

Also in adults, the link between non-verbal sensorimotor image schemas and language can be demonstrated. Participants associate verbs like *flee, point at, pull, push*, and *walk* with visuo-spatial scenes along the LEFT-RIGHT axis, while they associate verbs like *bomb, lift, smash, perch,* and *sink* with visuo-spatial scenes along the UP-DOWN axis (Richardson, Spivey, Edelman, & Naples, 2001). Abstract verbs are as likely as concrete verbs to be reliably classified into LEFT-RIGHT (e.g. *offend, argue with, warn, want, tempt*) or UP-DOWN spatial relations (e.g. *increase, hope, respect, succeed, own*). The association of image schemas with abstract words could be explained by early co-occurrences of concepts in sensorimotor experience, as assumed by the theory of primary metaphor (see 3.1.2 above). The verb *respect*, for example, is associated with the UP-DOWN image schema, perhaps because as children we look up to taller and wiser elders.

In another study the physical experience of standing was found to elicit the image schemas BALANCE, UP-DOWN, CENTRE-PERIPHERY, RESISTANCE, and LINKAGE (Gibbs, Beitel, Harrington, & Sanders, 1994). It can be shown that adults use the same image schemas in understanding the meaning of sentences describing physical and abstract uses of the verb *stand*, as in *The clock stands on the mantle* or *The part stands for the whole* (Gibbs et al., 1994).

It still remains to be identified what role image schemas play in the context of different representational levels of prior knowledge. The following levels of knowledge representation are distinguished (after Karmiloff-Smith, 1995; Mandler, 2004):

- *Procedural level*. Representations at this level have the form of sensory and motor procedures for analysing and responding to stimuli in the external environment. Knowledge at this level is not available to consciousness.
- *Image-schematic level*. These representations are more flexible without being tied to sensorimotor routines. Knowledge at this level is not yet available to consciousness, but can be processed cognitively (e.g. in priming experiments).
- *Explicit-analogue* and *explicit-verbal levels*. At these levels representations are accessible to consciousness. Explicit-analogue representations consist of spatial and kinaesthetic knowledge that is not accessible to verbal description. Explicit-verbal representations are accessible to verbal description (cf. the dual-coding approach of Paivio, 1990).

The relationship between these representational levels is shown in *Figure 3.4*, incorporated in the dynamic world model by Rasmussen (1986). Knowledge can be represented at several of these levels simultaneously and independently. Also, knowledge does not need to cover all of these levels. For instance, it is possible for us to tie our shoelaces (procedural level) without being able to give a verbal description of how we do so (explicit-verbal level). Each representation on a higher level is a more 'condensed' or 'compressed' version of the previous level (Karmiloff-Smith, 1995). Mandler proposes perceptual meaning analysis is a process of redescription that takes place while the infant is attending to events in the environment. Although PMA is mainly discussed as a process for building image schemas, it can also directly build new representations on the explicit-analogue and explicit-verbal levels (Mandler, 2004).

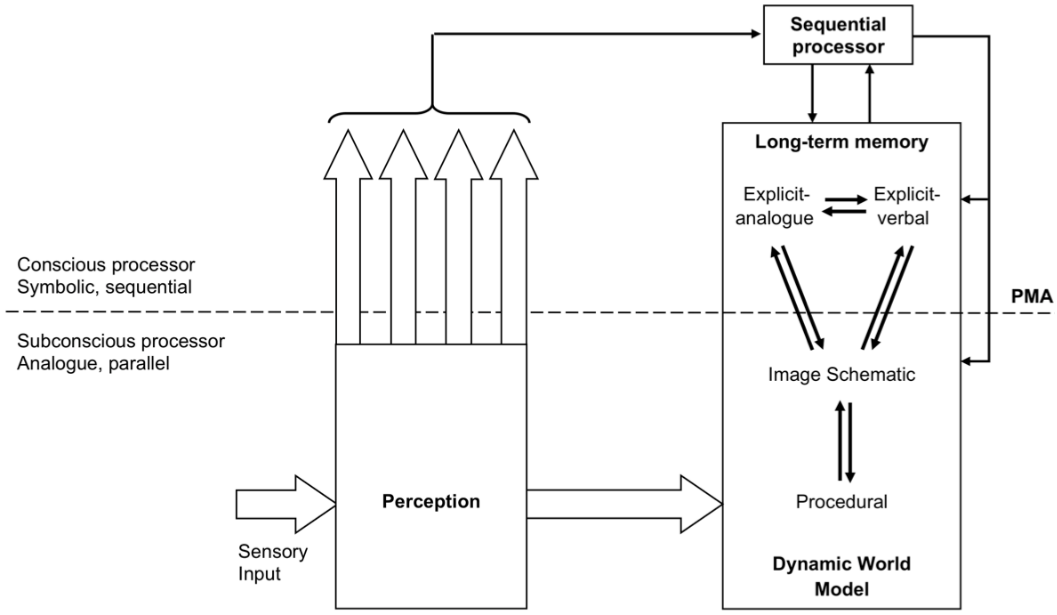


Figure 3.4. Image-schematic representations in the dynamic world model. Extract of the model human processor (Rasmussen, 1986) with added levels of representation (after Karmiloff-Smith, 1995) and the process of perceptual meaning analysis (PMA, Mandler, 2004).

To summarise, the developmental evidence suggests that the idea of image schemas as a sensorimotor form of representation is not only based on concepts resulting from detached philosophical speculation and subjective analysis. Image-schema theory is supported by recent findings in developmental cognitive psychology. The representational level of image schemas in *Figure 3.4* is beneath the threshold of consciousness. This leads the discussion to the second precondition of intuitive use – automatic processing of image schemas.

3.2.2 Subconscious Automatic Processing of Image Schemas

Most of the evidence for the automatic triggering of image schemas comes from studies of language understanding. First, these studies show that sentence processing evokes not only linguistic (or for that matter propositional) representations, but also evokes analogue schematic representations that resemble proposed image schemas. Second, these studies show that these analogue representations are evoked automatically and that they interfere with task-unrelated stimuli that are also present when processing the sentences.

These studies follow similar experimental procedures. Participants receive sentences that imply specific image schemas. For example, *John put the pencil in the cup* is used to evoke the UP-DOWN image schema and *John put the pencil in the drawer* is used to evoke the LEFT-RIGHT image schema (Stanfield & Zwaan, 2001). After hearing or reading these sentences, participants respond to a task unrelated to the image schema. For example, a picture appears after the sentence, and the task is to determine as fast as possible if the pictured item had been mentioned in the sentence or not. Importantly, in critical trials, the stimuli or responses include a hidden image-schema manipulation. In the critical trials, for example, the pictures contain objects that were mentioned in the sentence and the object's orientation is either consistent with the image schema implied by the sentence or inconsistent. Thus, the picture could show a pencil in vertical orientation after the presentation of *John put the pencil in the cup* (in consistent trials) or it could show a pencil in horizontal orientation after the same sentence (in inconsistent trials).

The hypothesis is that language understanding involves simulating the meaning of the sentence in which image schemas and procedural representations are involved (Barsalou, 1999; Glenberg & Kaschak, 2002). If image schemas are automatically activated, the hidden manipulation in these experiments should affect (e.g. facilitate) the responses in the otherwise unrelated task. Participants should respond faster to consistent pictures than to inconsistent pictures. If the sentences do not trigger image schemas, no differences in response times should occur between consistent and inconsistent pictures in the example.

Studies following this basic paradigm (Table 3.1) show that sentence comprehension automatically activates spatial representations that resemble the image schemas associated with the sentences. These representations interact with the processing of implicit non-verbal stimuli associated with the same image schemas. Most frequently a facilitation effect occurs: the image schema in the sentence primes spatial processing and accelerates the response in the choice reaction task. In some studies, however, the effect is an inhibition effect. Triggering the image schema verbally interferes with non-verbal spatial processing and responses are slower. In reviewing possible causes for these contrasting results, it was found that inhibition effects occur when there are no or only very short latencies between the verbal and non-verbal stimuli. Then, the triggering of the image schema has not finished before the spatial stimulus appears and both processes interfere with each other resulting in longer response times (cf. Bergen, Lindsay, Matlock, & Narayanan, 2007; Kaschak, Zwaan, Aveyard, & Yaxley, 2006; Richardson, Spivey, Barsalou, & McRae, 2003).

If it is assumed that image-schematic analogue representations are responsible for the effects obtained (which is claimed by some of the studies; others are not primarily concerned with representational formats), then many of the characteristics of image schemas are repeated by the results:

- Language comprehension is more than just accessing lexicon entries stored in long-term memory. Language understanding means to activate traces of perceptual and motor experience at other representational levels (cf. Glenberg & Kaschak, 2002; Zwaan & Taylor, 2006). This underlines the role of image schemas as bridging representations between sensorimotor experience and higher-level representations (see above, *Figure 3.4*).

Table 3.1. *Evidence for the Automatic Activation of Image Schemas*

Study	Task	Implicit Image-Schema Manipulation	Result
Richardson, Spivey, Barsalou, and McRae (2003; E1)	Shortly after listening to sentences participants indicated whether a symbol that shortly flashed on the screen was a circle or a square.	In critical sentences verbs were associated with LEFT-RIGHT (*argue*) or UP-DOWN (*sink*; cf. Richardson et al., 2001). Symbols flashed either above, below, to the left, or to the right of the centre of the screen.	Inhibition: Responses were slower for stimuli on the vertical axis after UP-DOWN verbs and on the horizontal axis after LEFT-RIGHT verbs. Results were comparable for abstract and concrete verbs.
Richardson, Spivey, Barsalou and McRae (2003; E2)	While hearing sentences pictures of the sentence's agent and patient appeared in sequence in the screen's centre. In test trials participants indicated whether they have seen two simultaneously presented pictures as illustrations of the same sentence.	In critical sentences verbs were associated with LEFT-RIGHT or UP-DOWN. In critical test trials the two pictures were simultaneously shown either in horizontal arrangement (side by side) or vertical arrangement (one above the other).	Facilitation: Yes-responses were faster when the arrangement of the two pictures in the test trial matched the verb's image schema. The effect was the same for concrete and abstract verbs.
Zwaan and Yaxley (2003)	Participants were asked to judge as quickly as possible whether two words are semantically related or not.	Words were presented one above the other. In the critical trials word pairs denoted parts of the same object, e.g. *branch-root*. UP-DOWN relations of word pairs were consistent (e.g. *branch* presented above *root*) or inconsistent with the usual spatial relation of their referents.	Facilitation: Responses were faster when spatial relations of the words and their referents were consistent. Effects of reading order could be ruled out.
Stanfield and Zwaan (2001)	After reading sentences participants determined whether a picture contained an item mentioned in the sentence or not.	In critical trials pictures contained objects mentioned in the sentence and were consistent with the UP-DOWN or LEFT-RIGHT orientation implied by the sentence or not (e.g. a pencil in vertical orientation after *John put the pencil in the cup* vs. *John put the pencil in the drawer*).	Facilitation: Responses were faster for pictures of objects whose orientation matched the orientation suggested by the prior sentence.
Zwaan, Madden, Yaxley, and Aveyard (2004)	After listening to sentences participants saw two pictures quickly appearing in sequence in the centre of the screen and indicated whether the two pictures were the same or not.	Critical sentences manipulated dynamic aspects of NEAR-FAR (e.g. *The shortstop hurled the softball at you* or *You hurled the softball at the shortstop*). On critical trials the object on both pictures was mentioned in the sentence (e.g. a softball). In the second picture the object was a bit larger or a bit smaller than in the first one, thus mimicking towards or away movements.	Facilitation: Responses were faster for picture sequences implying a movement that matched the movement implied by the sentence.

Table 3.1. *Evidence for the Automatic Activation of Image Schemas (Continued)*

Study	Task	Implicit Image-Schema Manipulation	Result
Kaschak et al. (2005)	Participants listened to sentences and made sensibility or grammaticality judgments.	Critical sentences manipulated dynamic aspects of UP-DOWN (e.g. *The confetti fell on the parade*) and NEAR-FAR (e.g. *The car approached you*). Simultaneous to their task, participants saw horizontal bars that moved up or down or spirals that moved towards or away.	Inhibition: Responses were slower for sentences implying motion that was consistent with the motion of the display.
Kaschak, Zwaan, Aveyard, and Yaxley (2006)	E1: Participants read sentences presented word-by-word and made sensibility judgments. E2: Participants listened to sentences and made sensibility judgments.	Critical sentences were the same in both experiments and manipulated auditory instantiations of UP-DOWN (e.g. *The jet pack roared into the sky*) and NEAR-FAR (e.g. *The dog barked loudly as it ran off across the prairie*). Simultaneous to their task, participants listened to auditory white noise stimuli conveying towards, away, upwards, and downwards motion.	Exp. 1: Inhibition: Responses were slower for sentences implying motion that was consistent with the auditory motion. Exp. 2: Facilitation: Responses were faster for consistent sentences
Glenberg and Kaschak (2002)	Participants read sentences and made sensibility judgements.	Critical sentences manipulated dynamic instantiations of NEAR-FAR describing physical transfers (e.g. *Courtney handed you the notebook, You handed Courtney the notebook*) or abstract transfers (e.g. *Liz told you the story, You told Liz the story*). Responses were given by pressing buttons that required either a towards-movement or an away-movement. The assignment of buttons to yes or no answers varied between conditions.	Facilitation: Responses were faster when the implied motion of the sentence matched the required motion for the response (e.g. away-sentences in the yes-is-far condition, and towards-sentences in the yes-is-near condition). The effect was consistent for physical and abstract sentences.
Glover et al. (2004)	Participants read a word presented on a screen and afterwards grasped a wooden block that could have one of three different sizes.	Critical words implied different sizes of objects (e.g. *apple* vs. *grape*), manipulating instantiations of the BIG-SMALL image schema. The dependant variable was grip aperture, i.e. the distance between thumb and index finger.	Facilitation: Participants had larger initial grip apertures after reading words representing larger objects than after reading words representing smaller objects.
Zwaan and Taylor (2006, E2)	Participants listened to sentences and made sensibility judgments.	Critical sentences were sensible and implied manual ROTATION, either clockwise (e.g. *Jenny screwed in the light bulb*) or counter clockwise (e.g. *Vincent dimmed the lights*). Responses were given by turning a knob clockwise for yes and counter clockwise for no (in one condition) or vice versa (in a second condition).	Facilitation: Responses were faster when the manual response matched the manual action implied by the sentence.

- The automatic activation of image schemas interferes with non-verbal image-schema activation in different modalities. Effects could be elicited in the visual and auditory modalities (Kaschak et al., 2005, 2006) as well as in the motor domain (Glover et al., 2004; Zwaan & Taylor, 2006). This underlines the multimodality of image-schematic representations.
- Automatic activation was obtained with rich stimuli (i.e. pictures with a high amount of detail) and with more schematic non-verbal stimuli alike. The Kaschak et al. (2005, 2006) studies, for example, show that it is sufficient to show moving horizontal bars to interfere with simulated UP-DOWN motions of rockets or to modulate the volume of white noise sounds to interfere with simulated NEAR-FAR movements of barking dogs. This underlines the schematicitiy of the underlying representational processes.
- Automatic activation was possible for static and dynamic instantiations of image schemas – whether it was the static spatial arrangement of words and their referents (e.g. Zwaan et al., 2004) or the dynamic movement of approaching cars and moving spirals (Kaschak et al., 2005). This underlines the dynamic flexibility of image schemas.
- Image schemas were automatically activated for concrete and abstract verbal stimuli. Although most of the available evidence involves words and sentences describing physical events, there is support for image-schema activation in the processing of abstract language as well (cf. Glenberg & Kaschak, 2002; Richardson et al., 2003). This underlines the use of image schemas for structuring abstract thought, as claimed by conceptual metaphor theory.

Although much can be learned from these studies, they all incorporated linguistic stimuli to activate image schemas. Only few studies are available that show how image schemas are automatically triggered by non-verbal stimuli. For example, the MOMENTUM image schema is activated in studies where participants see moving objects that are abruptly halted and are asked to indicate the last position of the object. The participant's answers are typically shifted in the direction of the represented motion. This effect is quite robust. It appears over large inter-stimulus intervals and with a range of stimuli including visual and auditory stimuli, rotatory, horizontal and vertical movements. When viewing vertical movements, for example, basic experiences like gravity are taken into account, so that the MOMENTUM effect is larger for downward movements than for upward movements (cf. Gibbs & Colston, 1995; Thornton & Hubbard, 2002). Other non-linguistic evidence comes from experiments on the perception of causal motion. It could be shown that 6 months old infants are responsive to the difference between an object starting to move on its own (SELF-MOTION) and an object starting to move after it was contacted by another moving object (COMPULSION) (Leslie, 1982; for the same data on adults see Michotte, 1963).

What can a user interface designer learn from these studies? First, image-schematic spatial representations can be activated automatically by linguistic, visual, auditory and motor stimuli. This means that user interface designs can be successful in activating image schemas in different modalities. Second, abstract concepts like *hoping*, *respecting* and *communicating* (Glenberg & Kaschak, 2002; Richardson et al., 2003) are structured by spatial image-schematic representations and thus can be represented by spatial means in the user interface (e.g. vertical sliders or horizontal paths of control movements). Third, UI elements that activate image schemas can be flexible in their design. They can be static or animated; they can be rich pictures or schematic ones.

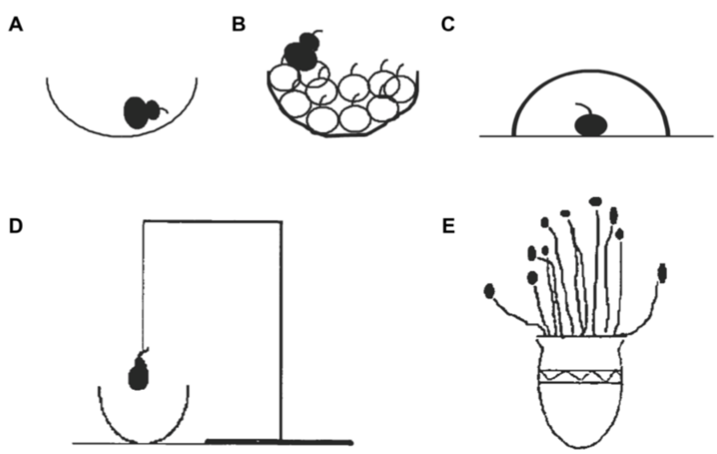


Figure 3.5. Various forms of CONTAINMENT (A, B, E) versus NON-CONTAINMENT (C, D) that are determined by geometry and function. Pictures taken from Feist (2000, p. 32, 184) and Coventry, Carmichael, and Garrod (1994, p. 290, 295).

While these studies have shown that different design features *can* automatically activate analogue image-schematic representations, the question remains *what features* are necessary to automatically activate a specific image-schema interpretation. One of the obvious features is topological relation. The pear in *Figure 3.5.A* is *in* the bowl, so the picture can be described as an instantiation of CONTAINMENT. Besides topology, other features determine the interpretation of CONTAINMENT relations. Although the content is (partly) outside of the container's boundaries in *Figure 3.5.B* and *Figure 3.5.E*, these scenes are interpreted as being about CONTAINMENT: the pear is still *in* the bowl and the flowers are *in* the vase. It follows that incomplete enclosure and being in the 'convex hull' of another object suffices to activate CONTAINMENT relations (Garrod, Ferrier, & Campbell, 1999; Talmy, 2000).

Enclosure or being-in-the-convex-hull-of-another-object are not sufficient to activate the CONTAINER image schema, as *Figure 3.5.C* and *Figure 3.5.D* illustrate. Here, instead of being *in* the bowl, the fruit is interpreted as being *under* or *above* the bowl, respectively. Although in *D* the pear is in a similar geometric relation to the bowl as in B, the UP-DOWN image schema is activated. This seeming peculiarity can be easily explained. In scenes *C* and *D* one of the entailments of the CONTAINER image schema is violated, i.e. that the location of the container determines the location of the content. This entailment is already active in 2.5-months-old infants (Hespos & Baillargeon, 2001) and can be demonstrated when adults interpret spatial scenes comparable to those in *Figure 3.5* (Garrod et al., 1999).

Consequently, it is not geometrical relation per se that activates image schemas; it rather is 'functional geometry' (Garrod et al., 1999; Coventry, Prat-Sala & Richards, 2001). Functional geometry also means that 'appropriate' objects are combined in the spatial scene. Bowls, for example, activate a CONTAINER image schema more strongly when they contain fruit than when they contain liquids. Jugs, in contrast, activate a CONTAINER image schema more strongly when they contain liquids than when they contain fruit (Coventry, Carmichael, & Garrod, 1994).

Although function and geometry may be the most salient factors determining the activation of image schemas, other influencing factors were found. Sometimes it is important how objects are labelled (Feist, 2000), whether absolute or relative frames of reference are used for spatial relations (Kelleher & Costello, 2005; Waller, Lippa, & Richardson, 2008), and even whether the related items are animate or inanimate (Feist, 2000).

For user interface design it follows that designers should use very abstract instantiations of image schemas to avoid the influence of, for example, functional factors. A second strategy is to actively explore whether adding subtle hints at the functionality of objects or providing verbal labels activates the desired image schemas more readily in the mind of the user.

3.2.3 Summary and Conclusion

The evidence shows that image schemas fit the preconditions of intuitive use quite well. The developmental evidence shows that image schemas are a form of prior knowledge that is derived from sensorimotor experience, but at the same time is much more abstract than mere recordings of the perceptual display. The developmental evidence also shows how image schemas can be a form of representation that bridges perception, action, and higher verbal and non-verbal representations. Several studies involving verbal and non-verbal stimuli confirm the hypothesis that image schemas are forms of implicit knowledge that is automatically activated. Implicit activations of image schemas facilitate or interfere with non-related tasks when these involve the activation of the same or of other image schemas. It has been shown that image schemas are activated in different modalities; and that static and dynamic stimuli, rich and schematic representations, as well as concrete or more abstract domains of knowledge activate them. A consequence of these findings is the broad area of application of image schemas in user interface design. However, what aspects of spatial scenes are involved in the activation of image schemas is still an open issue.

3.3 Addressing the Demand for Intuitive Use

The potential contributions of image schemas to meeting the demand for intuitive use are (1) that they facilitate the presentation and manipulation of abstract data by means of physical user interface elements; (2) that they support heterogeneous user groups; (3) that they support the design of hard- and software alike. The first promise directly derives from conceptual metaphor theory, the second derives from the sensorimotor grounding of image schemas, and the third derives from the abstractness of image schemas that can be instantiated in many different ways. Although plausible in theory, what is the evidence that supports these promises?

3.3.1 Supporting the Interaction with Abstract Data

Conceptual metaphor theory proposes that image schemas are the building blocks of abstract thought via direct metaphorical extensions grounded in repeated co-occurrences of domains of experience (the primary metaphor hypothesis) or via being recruited for metaphoric mappings between domains (the invariance hypothesis). Most of the evidence for conceptual metaphor has been accrued in cognitive linguistics. About 250 metaphorical extensions of image schemas are reported in the literature. Conceptual metaphor theory claims that linguistic metaphors are based on conceptual metaphors, but linguistic data do not suffice to claim the conceptual status of metaphors. Circular argumentation can only be escaped by conducting experiments that include non-verbal stimuli. Such experiments have been conducted and are most advanced in two areas: conceptual metaphors of time as a target domain and metaphorical extensions of the UP-DOWN image schema.

Conceptual Metaphors of Time

Talking about time involves using spatial expressions. One meets *at* noon (image schema LOCATION), a meeting lasts *from* two o'clock *to* four o'clock (PATH), and one can dance all *through* the night (CONTAINER). A specific metaphorical system is THE PASSAGE OF TIME IS MOVEMENT THROUGH SPACE. This metaphor comes in two versions: the ego-moving metaphor (e.g. *We're approaching Christmas, We passed the deadline, We are halfway through September*) and the time-moving metaphor (*Christmas is coming, The time for action has arrived, The deadline is approaching*). In both metaphors the PATH and the FRONT-BACK image schemas are apparent. In addition, the property of SELF-MOTION is assigned either to the observer (in the ego-moving version) or to the time itself (in the time-moving version). The two versions are mutually exclusive and are not readily mixed by speakers (Gentner, 2001).

In a series of experiments, participants saw spatial scenarios similar to those shown in *Figure 3.6*. Participants made true-false judgments on either ego-moving or object-moving spatial event frames. After several of these judgments participants read an ambiguous temporal sentence: *Next Wednesday's meeting has been moved forward two days* and were asked to indicate to which day the meeting had been rescheduled. If the sentence is interpreted within the ego-moving time frame, then *forward* is in the direction of the ego moving, i.e. the meeting then is on Friday. If the sentence is interpreted in the time-moving frame, however, *forward* is in the direction of motion of time and the meeting should now be on Monday. If space and time share relational structure that is transferred across domains the ego-moving spatial prime should facilitate an answer in terms of an ego-moving metaphor ('Friday'). Accordingly, object-moving scenarios should prime the time-moving metaphor and participants should be more likely to answer 'Monday' (Boroditsky, 2000).

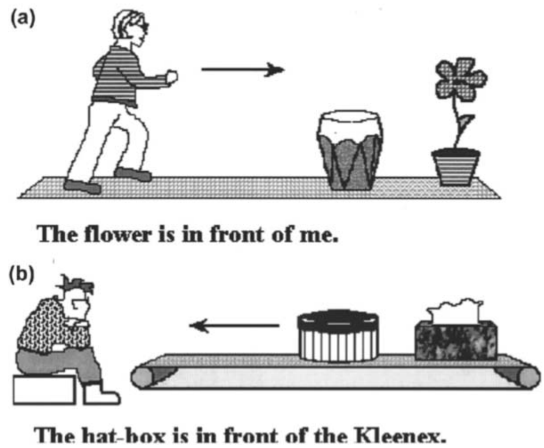


Figure 3.6. Sample scenarios used in the experiments by Boroditsky (2000, p.12, 13), A. Ego-moving spatial scenario, B. Object-moving spatial scenario

The results indicated that, when primed with the ego-moving frame of reference, 73% of the participants answered 'Friday' and 27% 'Monday'. When primed with the object-moving frame of reference only 31% answered 'Friday' and 69% 'Monday'. In a control group without priming, the answers were almost evenly split between Monday (46%) and Friday (54%). Thus, the availability of different spatial schemas affected how people thought about time.

A second more elaborate experiment showed that the reverse does not hold true. Participants gave prime-consistent answers when presented with spatial primes and temporal targets but not when presented with temporal primes and spatial targets. A third, more rigorous reaction-time based experiment, corroborated these findings. Again, spatial scenarios primed spatial and temporal stimuli while temporal scenarios only primed temporal stimuli, but not spatial ones (Boroditsky, 2000). These findings are consistent with the unidirectionality hypothesis discussed above (3.1.2). According to this hypothesis, the transfer of structure always is from the concrete source domain to the more abstract target domain and rarely the other way round.

The spatial priming effect on temporal ambiguous questions was confirmed in field studies (Boroditsky & Ramscar, 2002). People waiting in a line for lunch were more prone to give answers in the ego-moving frame the further they had been moving along the queue. At the airport, people that just flew in were more likely to use the ego-moving frame than people waiting for their departure. The latter were more likely to give ego-moving answers than people that just waited at the airport to pick someone up. These studies show that thinking about time is grounded in spatial experience – recent motion through space activates the ego-moving version of the metaphor.

Another study makes the case for the conceptual nature (in contrast to the linguistic nature) of the TIME IS SPACE metaphor more clearly, because it completely foregoes the involvement of language (Casasanto & Boroditsky, 2008). Participants saw lines growing on a computer screen. Lines varied in length and duration of growth. They grew until they reached their full extent and then disappeared. Participants were then randomly prompted to either reproduce the displacement or the duration of the line just seen. As expected, distance estimates were to a large degree influenced by the final length of the line and duration estimates were to a large extent influenced by the duration of the line growth. Regarding cross-domain mappings, the same pattern as in the studies of Boroditsky (2000) emerged. The actual distances covered by the lines influenced duration estimates. The greater the length of the line, the longer was the estimated duration (when in fact duration was held constant). In contrast, there was no effect of duration on the estimation of line length. This pattern of results remained stable across several variations of the experimental procedure, e.g. replacing the growing lines with a moving dot or presenting stationary lines for different durations (Casasanto & Boroditsky, 2008).

These results not only support the unidirectionality hypothesis of conceptual metaphor. They provide evidence that the metaphorical relationships between space and time in language also exist in our more basic representations of distance and duration in general. The results also suggest that abstract concepts like time are structured by representations of more physical experiences in perception and action.

Metaphorical Extensions of the Image Schema UP-DOWN

The linguistic evidence shows that the image schema UP-DOWN is metaphorically extended to a number of target domains. The most prevalent metaphors are GOOD IS UP – BAD IS DOWN, POWER IS UP – BEING SUBJECT TO POWER IS DOWN, MORE IS UP – LESS IS DOWN, and HAPPY IS UP – SAD IS DOWN.

The metaphor GOOD IS UP – BAD IS DOWN was studied by Meier and Robinson (2004). Participants evaluated words that were either in an UP or a DOWN position on a computer screen. Words either had positive or negative connotations (e.g. *hero, liar*). The results showed that positive words were evaluated more quickly when they were presented in the UP position than when they were presented in the DOWN position. Negative words were evaluated more quickly when they were presented in the DOWN position than when they were presented in the UP position.

In two follow-up experiments, the instantiations of UP-DOWN were separated from the evaluation of the words. In the first experiment, participants evaluated a word in the centre of the screen and then indicated which of two letters appeared either in the UP or DOWN position on the screen. After positive words response times were faster when target stimuli appeared in the UP position on the screen than when they appeared in the DOWN position (and vice versa for negative stimuli). In the second experiment, participants first indicated whether a stimulus appeared at the top or the bottom of the screen and then evaluated a word that was presented in the centre of the screen. Here, the spatial location did not prime the evaluation of words, e.g. UP did not activate GOOD. The unidirectionality hypothesis seems not to be confirmed here – activation seemed to spread from the target to the source domain, not vice versa as in the studies above.

The metaphor POWER IS UP – BEING SUBJECT TO POWER IS DOWN was studied by Schubert (2005). Participants saw two power-related words at a time on a computer screen (e.g. *master – servant*) and either decided which one denoted a powerful or which one denoted a powerless social group. When words were presented in metaphor-consistent arrangements (e.g. *master* above *servant*), reaction times were shorter than when the words were presented in metaphor-inconsistent arrangements (e.g. *servant* above *master*). Two subsequent studies separated this effect into a perceptive and a response component and showed that the effect independently occurs in both of these components.

In other experiments it was shown that not only reaction times but also powerful / powerless judgements themselves are influenced by the position on screen (Schubert, 2005, experiment 6) and that the length of vertical, but not horizontal, lines can prime the degree of power attributed to leaders of a company when lines are displayed in an organisation chart (Giessner & Schubert, 2007).

What happens when the two metaphors GOOD IS UP and POWER IS UP make contradicting predictions? For instance, the words *enemy* or *dictator* are denoting powerful and at the same time negatively evaluated groups of people. Hence, each of the two metaphors predicts a different location: POWER IS UP and BAD IS DOWN. The evidence shows that the results depend on the task given to participants (Schubert, 2005, experiment 5). When participants judged whether groups of people were powerful or powerless, then response times were sensitive to the power metaphor. When participants judged whether groups of people were good and bad, then response times were sensitive to the valence metaphor.

The processing of UP-DOWN metaphors also interacts with UP-DOWN gesturing (Casasanto & Lozano, 2006) and motor action in general (Casasanto & Lozano, 2007). In one of these experiments participants looked at a computer screen on which abstract words appeared either in blue or red font. These words belonged to the target domains of several UP-DOWN metaphors (e.g. *wealthy, poor, virtuous, evil, joy, disgust*). Stacked next to the screen, on the left and right side, there were three boxes. The top box was red and the bottom box was blue (or vice versa in another condition). The middle box was white and contained marbles. Participants were instructed to only respond to the colour of the words. Colours indicated into which box the marbles should be transferred, thus inducing upwards or downwards movement of both hands simultaneously. Although participants were not instructed to read or to remember the task-irrelevant words on the screen, a surprise old / new recognition memory task was presented at the end of the experiment.

Participants moved the marbles faster when the movement was consistent with the image schema associated with the abstract word (e.g. *wealthy* plus UPward movements of marbles) than when the movement was inconsistent (e.g. *evil* plus an UPward movement). There was also a strong effect on recognition memory. Participants correctly recognised 94% of the words that were incidentally encoded during consistent marble movements, but were at chance level (54%) recognizing words encoded during inconsistent marble movements. As a final observation, marbles were lost more often during image-schema inconsistent movements (Casasanto & Lozano, 2007).

The Casasanto and Lozano studies (2006, 2007) show that the unidirectionality hypothesis is difficult to hold in practice. While sometimes findings are compatible with the unidirectionality hypothesis (Boroditsky, 2000; Casasanto & Boroditsky, 2008) and sometimes not (Meier & Robinson, 2004), the marble moving studies show that both ways are possible: the target domain (abstract words) influences the source domain (marble movement) and the source domain influences the target domain (word memory). It rather seems that, once neural associations between concepts are established, activation can spread both ways.

In some studies the use of verbal stimuli is excluded altogether. In one study investigating the GOOD IS UP – BAD IS DOWN metaphor, participants viewed pictures that were either positive (e.g. a picture of a smiling face) or negative (e.g. a picture of a dead animal). Pictures appeared randomly on various horizontal and vertical locations on a computer screen. In a subsequent testing phase, each picture appeared in the centre of the screen and participants were instructed to use the mouse to move the picture back to the location where it appeared previously. The results indicated that, compared to the original vertical positions of the stimuli, the memorised position of positive stimuli was biased upwards while the memorised position of negative stimuli was biased downwards. No evidence was found for an affective bias in the horizontal direction (Crawford, Margolies, Drake, & Murphy, 2006).

Conclusions: Conceptual Metaphor in User Interface Design

More evidence on the psychological reality of conceptual metaphor is available than there is space to review here. They include a broader range of image schemas as source domains as well as occurrences in different modalities – vision, gesture, and hearing. Studies on conceptual metaphors of the NEAR-FAR image schema are discussed in chapter 6. Evidence for the association of affective states with the image schemas UP-DOWN and BRIGHT-DARK can be found in Meier and Robinson

(2006); Meier, Robinson, Crawford, and Ahlvers (2007); and Weger, Meier, Robinson, and Inhoff (2007). For conceptual metaphor in gesture see Cienki and Müller (2008), McNeill (1992, 2005), and Sweetser (1998). For research on metaphors in pitch perception, e.g. HIGH PITCH IS UP – LOW PITCH IS DOWN, see Casasanto, Phillips, and Boroditsky (2003) as well as Rusconi, Kwan, Giordano, Umiltà, and Butterworth (2006). Neurological evidence for conceptual metaphor is discussed in Kemmerer (2005) and Rohrer (2005).

The studies presented above go beyond pure linguistic evidence and demonstrate that spatial and sensorimotor experiences influence and interact with thinking about abstract concepts. Many of these studies involve the implicit activation of metaphorical associations while solving unrelated tasks. Hence, the evidence seems to support the claim that conceptual metaphor operates automatically and beneath conscious awareness.

Most of the studies above are laboratory based and their relevance to user interface design is difficult to determine. Unfortunately, conceptual metaphor theory so far has only been a guiding philosophy for user interface designers and research is missing that comes close to the methodological rigour of the studies conducted in experimental psychology and at the same time makes relevant the importance of conceptual metaphor for design for intuitive use. Despite their limitations, user interface designers can learn the following from this research:

The studies have shown that physical properties like spatial positions on a display or physical movements performed by the user can influence the thinking about abstract domains. Or vice versa: abstract domains can activate image-schemas that interfere with or facilitate the interpretation of spatial displays or physical movements. Although this seems to violate the unidirectionality hypothesis, the exact conditions and dependencies that explain the violation in some cases and the strict unidirectionality of other findings are not clear. Further research is needed to clarify this issue.

The main message for user interface design, however, is that abstract domains of knowledge can be effectively communicated via input and output devices that necessarily are physical – and that the associations between the concrete and the abstract are not arbitrary but follow systematic patterns.

Often, in real-world contexts, metaphors will contradict each other. In charting the number of enemies in a computer game, for example, the metaphors MORE IS UP and GOOD IS UP make different predictions. More enemies are bad – should the chart go up or down? Here, the result of Schubert (2005) gives an indication. The appropriate metaphor could be determined by the task. If the goal is to convey information about the quantity of enemy's forces, the metaphor MORE IS UP will be more appropriate. If the goal is to convey information about the emotional impact of the enemy's forces, the metaphor GOOD IS UP is more appropriate and a virtual 'situation-valence meter' will indicate falling values.

If user interface designers want to generalise from the available evidence that the more than 250 image-schematic metaphors purported by cognitive linguists are conceptual metaphors, then a vast amount of potential guidelines for physical-to-abstract mappings is in place and needs only to be tapped. What is more, designers do not have to wait for linguists to find metaphors that suit their needs. The invariance hypothesis means that designers can analyse abstract domains directly to extract image-schematic content. At the current state of research, however, it is advisable to see conceptual metaphor theory as a theoretical and heuristic guideline rather than an empirically proven framework. The unidirectionality hypothesis and the invariance hypothesis, for example, are still what their name suggests: hypotheses – inspired by linguistic findings, confirmed by empirical evidence from some domains, but awaiting further investigation in other domains.

3.3.2 Supporting Heterogeneous User Groups

Image schemas are claimed to derive from basic sensorimotor experiences in the world and their metaphorical extensions derive from correlations with subjective experiences or judgments. This grounding in basic experiences could mean that image schemas and their metaphorical extensions are universal and hence could support user interface design for very heterogeneous target groups. As with most of the evidence on image schemas and conceptual metaphors, the evidence for their universality stems mainly from linguistic studies.

Universality of Image Schemas

Linguistic studies confirm the universality of image schemas. One of the major finding is that although linguistic terms can differ across languages, the concepts underlying them do not. Much research has been done on closed-class linguistic elements, i.e. groups of expressions that usually contain a relatively small number of items and to which no new items can normally be added. Typical closed-class expressions that are found in many languages are adpositions (prepositions and postpositions), determiners (e.g. articles, quantifiers, demonstratives), conjunctions (e.g. *and*, *nor*, *but*), and pronouns. Much work has been done on spatial closed-class elements. Talmy (1983, 2005), for example, looked on spatial expressions in languages like Atsugewi (a Californian Indian Language), English, Finnish, Hebrew, Japanese, and Korean and found that only a limited set of object properties and spatial relations are coded into these elements. Although Talmy does not refer to the notion of image schemas, the categories and elements found in his analysis are very similar to image schemas, e.g. CONTAINER, PATH, SURFACE, NEAR-FAR, CENTRE-PERIPHERY, and so on. Talmy (2005) is confident that his inventory represents the main categories that can be found most often across spatial expressions in languages.

In an empirical study on the naming of spatial relations, line drawings were presented to speakers of nine genetically different languages (Levinson & Meira, 2003). Line drawings depicted spatial scenes such as a ring on a finger, a cloud over a hill, or a rope around a tree stump. Participants were asked *Where is the ring/cloud/rope/...?* A direct comparison at the surface level of linguistic terms showed that different expressions are associated with the same spatial scenes. This is comparable to the finding that while English speakers use the same preposition in describing spatial scenes where something is *on the wall* or *on the table*, Germans will distinguish these scenes by using different prepositions: *an der Wand* or *auf dem Tisch*. To arrive at the deeper structure of spatial relation terms, pictures that elicited the same responses were given high similarity scores and pictures that elicited different responses were given low similarity scores. These scores were then subjected to a multidimensional scaling (MDS) statistical analysis. The MDS revealed a solution in which pictures cluster around a relatively small number of core concepts that recur across languages. These clusters can be described as containment, attachment, superadjacency-subadjacency, and proximity (cf. Kemmerer, 2006) – corresponding to the image schemas CONTAINER, CONTACT, UP-DOWN, and NEAR-FAR. The MDS solution remained stable when data of further languages were added (Levinson & Meira, 2003).

Levinson and Meira (2003) also looked at how languages fractionate spatial concepts. The most general spatial concept mirrors the image schema LOCATION (corresponding to the English preposition *at*). If a language differentiates two spatial categories, it is CONTAINMENT and the rest of the LOCATION category. If a language differentiates more spatial categories, the LOCATION category splits into UP+CONTACT (corresponding to the English preposition *on*), DOWN, and NEAR. Later, UP+CONTACT are separated. Note that the systematic division of spatial terms across languages reflects the order in which children acquire locative terms (Johnston, 1988). It seems then, that some languages stop at some point in the acquisition of spatial relation concepts leaving further distinctions unreflected in their closed-class spatial relation expressions.

Other studies confirm that languages share a number of underlying structuring concepts. Munnich, Landau, and Dosher (2001) found that although English, Japanese, and Korean differ in the prepositions used to describe spatial relations, visual memory for location is structured very similarly across languages. Location memory was highly dependent on the UP-DOWN, LEFT-RIGHT, and CONTACT relations to referent objects (as opposed to e.g. oblique relations). Segerer (2008) investigated 72 languages from all parts of Africa in which adjectives belong to the group of closed-class linguistic elements. He found that the distinctions BIG-SMALL and BRIGHT-DARK are among the most popular distinctions, along with the more abstract concepts of *new-old* and *good-bad*. Wierzbicka (1996) develops a 'Natural Semantic Metalanguage' (NSL) that tries to break the semantics of words down to a set of about 60 semantic primitives believed to be atomic, primitive meanings present in all human languages. The NSL hypothesis was tested extensively against nine languages: Polish, Mandarin, Malay, Lao, Spanish, Korean, Mbula (Austronesian), Cree (Algonquian), and Yankunytjatjara

(Australian Aboriginal). Wierzbicka (1996, p. 22) claims that NSL reflects a "language-like innate conceptual system", and could in principle be used as the sole metalanguage for semantic definitions. The set of her primitive concepts show great overlap with the set of image schemas originally developed by Johnson (1987).

All these studies point to a certain universality of image schemas across different languages and cultures. However, as with image schemas in general, there is a lack of cross-cultural studies that demonstrate the psychological reality of image schemas using non-verbal material. Nevertheless, the great promises of image-schema theory clearly warrant such studies.

Universality of Conceptual Metaphor

In conceptual metaphor theory, broad emphasis is put on the experiential grounding of metaphors. Complex metaphors are said to be compounds of more basic metaphors derived from early and repeated experiences with the environment: "When the embodied experiences in the world are universal, then the corresponding primary metaphors are universally acquired" (Lakoff & Johnson, 1999, p.56). This view has been challenged by many who believe that culture also plays an important role in the formation of concepts, and thus metaphors cannot be universal (Correa-Beningfield, Kristiansen, Navarro-Ferrando & Vandeloise, 2005; Kimmel, 2005; Murphy, 1996; Rakova, 2002; Schmitt, 2001).

It turns out that both views are qualified: conceptual metaphors are universal and influenced by culture. Let's begin with their universality. The image schema STRAIGHT, for example, is used in languages as different as English, Hungarian, Japanese, and Russian in basically the same metaphors, e.g. MAXIMALLY INFORMATIVE SPEECH IS STRAIGHT as in *straight talk, a straight answer*, or MORAL IS STRAIGHT as in *to follow the straight path, he is crooked* (Cienki, 1998). Although these metaphors are the same across languages, their content may be evaluated differently. While *straight talk* is a good thing in American culture, it is regarded impolite according to Japanese cultural standards.

In a comprehensive survey of the universality vs. cultural issue, Kövecses (2005) shows that metaphors about emotions, event structure, time, and the self are consistent across languages of different families (such as English, Chinese, Hungarian, Japanese, Polish, Wolof, and Zulu). Only minor variations occur, e.g. in the word forms that are used (e.g. nouns instead of verbs) or in the different salience of parts of the metaphor. The emotion of anger, for example, is cross-linguistically conceived as ANGER IS HEATED SUBSTANCE IN A PRESSURISED CONTAINER as in *You make my blood boil. Simmer down! He was bursting with anger.* The shared physiology of the emotions is what makes this metaphor universal. There is variation in the details, however. In Chinese, the HEAT notion is de-emphasised, while in Wolof the notion of PRESSURE is absent. In Japanese the CONTAINER for anger is the belly, while in Zulu it is the heart, and in English the whole body. Very often variation is dependent on cultural models of the body's functioning, e.g. the notion of *qi* in China or the cultural significance of the belly area, *hara,* in Japan.

The idea of largely language-independent mechanisms of metaphor production is supported by a study of Neumann (2001), who reports of 106 analogous metaphors in German and Japanese; this body of comparable metaphors having been extended during the last years to include more than 300 (C. Neumann, personal communication, April 20, 2009).

While image-schematic metaphors seem closer to the universal end of the continuum, compound metaphors are more likely to be formed by culture. Although compound metaphors are, according to theory, built from primary metaphors, it is highly dependent on culture how many and which primary metaphors are assembled and which are more salient than others (Kövecses, 2005).

Metaphors can also vary between subcultures, as different expressions in different ethnic and/or religious groups show. Also, region has an influence. For example, Dutch and its derivate, Afrikaans, differ in the amount of mountains and animals that are present in their figurative languages (Kövecses, 2005). Probably due to historical and geographical reasons, PATH metaphors in English are more often about ships and sailing than those in French (Boers, 1999).

Metaphoric expressions can also change over short time frames within a culture. Boers (1999) investigated the editorials of the weekly issues of *The Economist* over a time period of 10 years. He found that the use of metaphors drawing on the source domain of HEALTH (*a healthy economic climate, symptoms of a corporate disease*) almost doubled during the winter period, i.e. from December to March, compared to the rest of the year. Indeed, winter is the time when the number of colds, bronchitis, influenza, pneumonia, etc. is increased, at least in the Northern hemisphere (cf. CDC, 2008). Boers argues that if people are more often confronted with health problems in winter, they are also more likely to use this bodily experience as a source domain for metaphorical mappings in that season. Other metaphors like THE ECONOMY IS A RACE/WARFARE/MACHINERY did not show the same time-dependant pattern, thus excluding the possibility that the findings are a result of an overall increase of metaphorical activity in the winter season.

Again, cross-cultural studies excluding linguistic material are harder to find. Variations in the TIME IS SPACE metaphor occur in speakers of English, Indonesian, Greek, and Spanish. When talking about amounts of time, distance or quantity metaphors can be used, e.g. *a long time* vs. *much time.* Casasanto et al. (2004) could show that variations of the line-growing task (Casasanto & Boroditsky 2008, see above), here line-growing versus container-filling, had different effects in speakers of different languages and that these effects were consistent with the preference for either distance or quantity linguistic metaphors in the language they spoke. People speaking 'distance' languages like English and

Indonesian produced larger effects in the task of growing lines than in that of filling containers. People speaking 'quantity' languages like Greek and Spanish produced larger effects in the task of filling containers than in the task of growing lines.

Further evidence on language variations in metaphor was suggested by Boroditsky (2001) who investigated the understanding of time by speakers of Mandarin and English. In contrast to English, which only knows horizontal metaphors of time (i.e. THE FUTURE IS IN FRONT – THE PAST IS IN THE BACK), Mandarin also has vertical metaphors (i.e. THE PAST IS UP – THE FUTURE IS DOWN). Accordingly, it was found that English speakers' conceptions of time can be primed by horizontal stimuli but not by vertical stimuli. Mandarin speakers' conceptions of time were primed by both, horizontal and vertical stimuli. However, as subsequent studies failed to replicate these effects (Chen, 2007; January & Kako, 2007), the stability of intercultural differences is less clear.

The conclusion is that metaphors are grounded in experience. This experience can be about basic correlations in sensorimotor experience – as probably is the case for image-schematic primary metaphors. Experience can also mean linguistic and social-cultural experience that forms the use of more complex metaphors.

The consequence for user interface design is that three different approaches can be considered. The first approach is to empirically test hypothetical metaphors across cultures. From the results guidelines can be derived as to which metaphors work in which culture. This process of research is needed, but requires much effort: studies need to be carefully designed and administered to members of different cultures and linguistic families. As only a small set of metaphors over time can be included in such studies, the whole process will take a long time and will therefore be impractical to user interface designers.

A second, more practical, approach could be to take informed decisions based on theoretical considerations. If a good line of argumentation can be created for the grounding of specific metaphors in basic experience and many people share these basic experiences across cultures, then universality may be assumed. This could be the case for many primary metaphors like MORE IS UP, DIFFICULT IS HEAVY, or SIMILAR IS NEAR. However, cross-cultural tests of these metaphors will still be necessary. To avoid minor cultural variation on otherwise equal image-schematic content, metaphorical instantiations could be reduced in detail, i.e. designers could use abstract visualisations of the metaphor. This would include using an abstract instantiation of a CONTAINER image schema (e.g. a two-dimensional area enclosed by a line) instead of instantiations that are highly culture specific (e.g. pictures of paper baskets or wineskins).

The third approach is not unknown to designers familiar with user-centred design and is summarised in the slogan *Know thy user!* (cf. Gould & Lewis, 1985). With conceptual metaphor theory as a background, it still remains necessary to study users, their tasks and mental models, to discern what conceptual metaphors they use in their language and thought and to determine what image schemas are underlying these metaphors.

3.3.3 Applicability to Hardware and Software User Interfaces

One trend that fuels the demand for intuitive use is the trend towards software user interfaces that allow rich and flexible options for displaying information but have reduced input capabilities. The current guidance for designing intuitive use sometimes is not suitable for being applied in soft- and hardware alike. Real affordance, for example, can by definition only exist in hardware and any more complex user interface metaphors are more easily incorporated in software than in hardware.

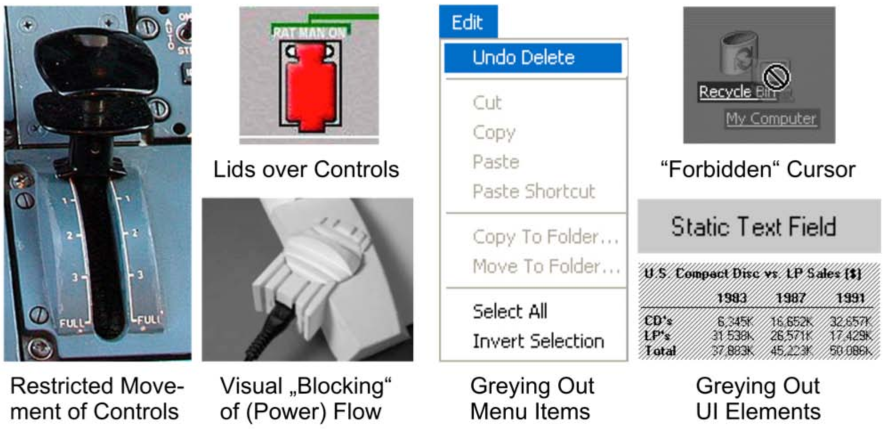


Figure 3.7. Different instances of BLOCKAGE in hardware (left Airbus A320 and Videodeck, Djajadiningrat, Wensveen, Frens, & Overbeecke, 2004, p. 300) and software user interfaces (right, Microsoft Windows)

Image schemas and their metaphorical extensions, however, should be equally suitable for the design of hardware and software. This follows from the studies discussed in 3.2 and 3.3 that have used instantiations of image schemas in the visual, auditory, and motor modalities as well as in language. Indeed, examples from real user interfaces indicate that image schemas can have a large variety of instantiations. Even FORCE image schemas that may be assumed to only be applicable to hardware can also be found in software user interfaces (see *Figure 3.7* for an example of the FORCE image schema BLOCKAGE; other FORCE image schemas include RESISTANCE, COMPULSION and ENABLEMENT). In hardware, BLOCKAGE can be conveyed via the impossibility to move levers or hand gears. BLOCKAGE is also present when controls are covered by lids that must be lifted or glass panes that need to be broken before the control can be used. Another example is a rotary power switch that indicates the ENABLEMENT or BLOCKAGE of electricity flow by aligning ribs on the switch with the ribs on the transformer's body or putting them at a right angle (see *Figure 3.7*). In software, BLOCKAGE is conveyed via greying out menu entries, input fields, or whole areas of the screen. Other instances of BLOCKAGE are error messages, login screens asking for passwords, user interface elements that do not react to mouse clicks or other input from the user, or show a "forbidden" cursor if a user action is not allowed (*Figure 3.7*). The examples show that image schemas can in principle be instantiated in software and hardware alike. However, a more systematic study needs to explore a further range of image schemas.

3.4 Summary and Conclusion

This chapter discussed the theoretical foundations of image schemas and their metaphorical extensions as well as the empirical evidence for them. Image schemas are recurring patterns of sensorimotor experience that form analogue and multimodal representations in the mind. Image schemas form experiential gestalts in the sense that they have internal structures that cannot be broken apart without destroying the specific entailments image schemas have. With these characteristics they form important building blocks of thought. In user interfaces they can be used to represent physical structures by simple physical-to-physical relationships.

Conceptual metaphor theory describes how image schemas are recruited to help structure more abstract domains. This can happen via primary metaphors that arise when experiences structured by image schemas co-occur with other subjective experiences or judgments. As about 250 of such primary metaphors have been published in the literature so far, primary metaphors can be potential guidelines for design, prescribing physical-to-abstract mappings in user interfaces.

The invariance hypothesis proposes that image schemas are involved in more complex metaphorical mappings by providing the basic structure that is mapped from the source to the target domain. Although more speculative than the primary metaphor approach, it resonates well with recent accounts of structure mapping in the explanation of metaphor and analogy (Bowdle & Gentner, 2005; Gentner, 1983). Applied to user interface design, the invariance hypothesis implies that image schemas are involved in forming the underlying structure of understanding more complex task domains and users' mental models. Following the basic principle of designing for intuitive use (chapter 1), image schemas can then be used for prescribing the structure of future user interfaces.

Many of the theoretical foundations of image schemas were derived from philosophical enquiry and linguistic investigations, and their claims about the conceptual nature of image schemas have been confirmed in psychological studies. Studies on the acquisition and representation of image schemas in pre-verbal infants give a plausible account of how image schemas are derived from sensorimotor input during the early months of life and how they form the bridge to higher-level concepts and language. These ideas are in line with more recent accounts of concept forming (Barsalou, 1999) and mental scaffolding of higher-level concepts (Williams, Huang, & Bargh, 2009).

For applying image schemas in the design for intuitive use this means that image schemas fulfil the defined precondition of intuitive use, because they are well anchored as a form of prior knowledge that is processed beneath conscious awareness. A number of studies have shown that image schemas are automatically triggered. However, most of this research was concerned with the sensorimotor implications of language understanding. It is less clear which non-linguistic factors play a role in activating image schemas, as not only geometrical relations are important, but also functional geometry and context. For transferring these findings to user interface design, further research needs to determine how specific layouts and appearances of user interface elements affect the activation of image schemas in the user's mind.

With regard to meeting the demand for intuitive use, the greatest promise of image schemas lies in the design for abstract data. The results of psychological experiments suggest that the claims of conceptual metaphor theory about the conceptual foundation of linguistic metaphor are warranted. A range of experimental paradigms has been employed to show that image-schematic metaphorical mappings are automatically activated when thinking about abstract stimuli. Experiments included verbal and non-verbal stimuli presented in the visual and motor domain. Results were obtained that – translated into outcomes of intuitive use – mean that applying metaphor could lead to more effective and mentally efficient interaction. The third outcome of intuitive use, satisfaction, was not measured in previous studies. Whether these metaphors are usefully applicable in user interfaces is an open question that is addressed in the following chapters.

Another promise is that image schemas meet the demand of designing for heterogeneous user groups. Again, the evidence mostly derives from linguistic studies. Although languages differ on the surface level of vocabulary, there is some evidence that image schemas form the underlying structure of at least closed-class linguistic elements like prepositions or basic adjectives. The evidence on conceptual metaphor suggests that at least primary metaphors are cross-culturally valid (with minor variations), and there are greater variations in more complex metaphors. As most of this evidence is linguistic and controlled studies manipulating the conceptual content of these metaphors are rare and contradictory, much research still needs to be undertaken regarding cross-cultural issues. The theory suggests universality; therefore, designers are encouraged to use image schemas in user interface design. At the present status of empirical evidence, however, designers are on the safe side when they keep following standard human-centred design methodologies that require user research at the start of the project and iterative testing with target users in later phases.

From the theoretical and practise perspective, there is good reason to assume that image schemas can be applied to the design of hard- and software user interfaces alike, thus being able to meet a third demand for intuitive use.

Taking a look back it can be seen what has been achieved so far. A definition of intuitive use is available and the consequences of intuitive use can be measured in terms of effectiveness, mental efficiency and satisfaction (chapter 2). The theoretical foundation of image schemas is established, as well as how image schemas meet the preconditions of and the demand for intuitive use (this chapter). This means that the basic theoretical prerequisites to evaluate image schemas as a new design tool for intuitive use are given. The next chapter reviews previous applications of image schemas in user interface design and derives the scope of the empirical research that is described in chapters 5 to 8.

4 Image Schemas and User Interface Design

It is time to draw the strings of the previous chapters together. First, intuitive use results from a match between the user's mental model and the represented model at the user interface (chapter 1). The invariance hypothesis suggests that image schemas form a significant part of the structure of the mental model when thinking about abstract domains (chapter 3). Therefore, analysing the image schemas in the tasks and mental models of the users and applying them to designing user interfaces should enhance intuitive use.

Second and more specifically, intuitive use is defined as the subconscious application of prior knowledge that leads to effective, mentally efficient, and satisfying interaction (chapter 2). Image schemas and their metaphorical extensions are forms of prior knowledge that are applied subconsciously, because they are learned early, experienced repeatedly, and can be evoked automatically (chapter 3). Therefore evoking image schemas and their metaphorical extensions through user interface design should support effective, mentally efficient, and satisfying interaction.

Third, image schemas can meet several demands for intuitive use (chapter 1 and 3). They are assumed to be suitable for heterogeneous user groups, flexibly applicable to hard and software, and enable physical-to-abstract mappings via primary metaphors.

Many open questions are posed by these conclusions. The strategy in this work is to focus on two specific aspects. The first extends the previous cognitive science research on the psychological reality of image-schematic metaphors to user interface design. In four controlled experiments it is investigated whether user interfaces congruent with image-schematic metaphors are more effective, mentally efficient, and satisfying to use than user interfaces that are not congruent with the metaphors (chapters 5 and 6).

The second focus is on investigating the implications of the invariance hypothesis. The goal is to learn whether it is possible to apply image schemas to the analysis of users' tasks and mental models and transfer the findings to the design of user interfaces (chapter 7 and 8). To prepare the ground, the next sections in this chapter first look at previous work that applied image schemas in a user-centred design. Then an image-schema vocabulary is presented that is used in the design studies. The final section draws together the research questions addressed in the empirical work as well as those questions that are left to future research.

4.1 Previous Work

It is useful to structure the previous work on image schemas in user interface design along the activities in a standard human-centred design process (ISO 13407, 1999; *Figure 4.1*). In the first phase, the context of use is analysed and specified. The results are fed into the second phase, requirements specification. In the third phase, design solutions are produced. These are evaluated in the fourth phase. The process is iterative so that it can be repeated until the product meets the specified goals.

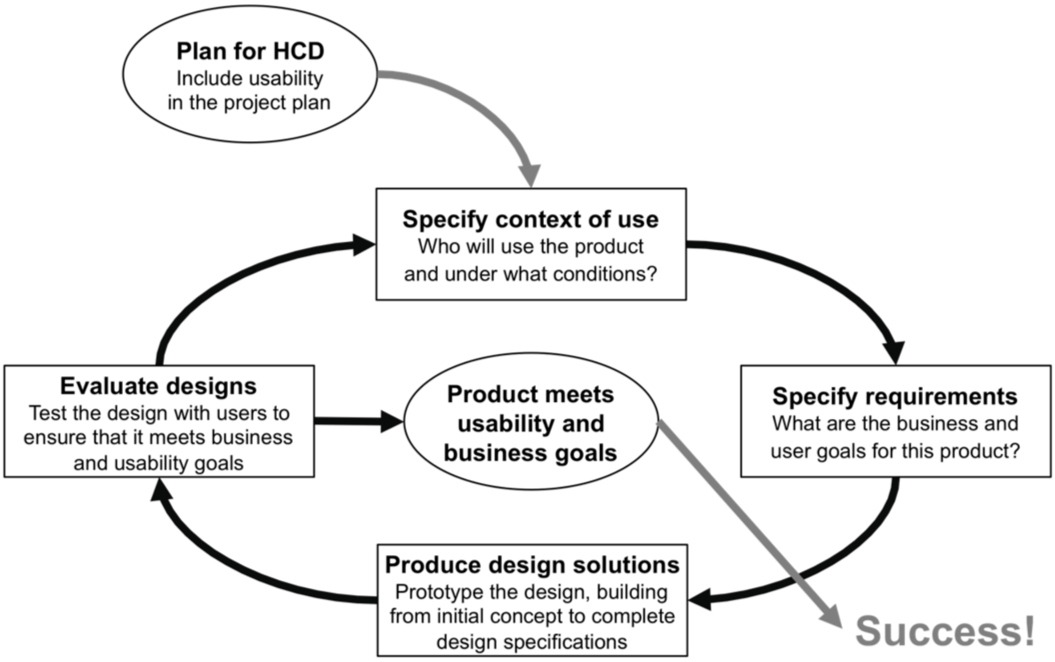


Figure 4.1. The human-centred design (HCD) process according to ISO 13407 (ISO, 1999).

4.1.1 Image Schemas in the Analysis and Requirement Phases

In the first phase of the human-centred design cycle, the context of use typically is analysed in situ. Characteristics of the task to be solved (including user goals), the current technological support, the characteristics of the target user group, and the general organisational context are analysed. In the first phase of the process, image schemas can be applied in understanding and specifying the context of use. Here, image schemas can describe the structure of the task, the mental model of the user, and the current user interface design.

In the second phase, user and organisational requirements are derived and specified from the results of the context-of-use analysis. Here, image-schema analysis can inform the specification of requirements in the sense that the abstract structure of the task and the mental model described in image-schema vocabulary are used as a prescription for design.

Previous studies show, for example, that image schemas can be extracted from users' utterances, thus revealing parts of their mental models. Maglio and Matlock (1999) analysed users' mental models of the World Wide Web (WWW) using image schemas. Although the WWW is usually described as a COLLECTION of web sites (LOCATIONS) that are connected via LINKS, the users' language about the web reveals many instances of the image schemas SELF-MOTION, CONTAINER, and PATH. The use of PATH metaphors even increased, when the participants were more experienced with using the WWW.

In other studies, the mental models of people were examined when navigating in airports (Raubal, 1997; Raubal & Worboys, 1999). Image schemas were extracted from the utterances made by users finding their way through simulations of these airports. Image schemas could be extracted from almost all utterances of the participants, and suggestions for the redesign of the airport navigation system could be derived from the results.

With the prospect of building interactive environments for learning musical concepts, Bakker, Antle, and van den Hoven (2009) extracted image-schematic mappings from the body movements of 7 to 9 year old children. The children were asked to move their body or an object to enact musical samples in each of which one of eight sound concepts was changing (i.e. volume, tempo, pitch, tone duration, timbre, rhythm, harmony, and articulation). The results identified 27 metaphorical mappings whereby many target domains mapped to multiple image schemas, e.g. LOUD SOUNDS ARE BIG – SOFT SOUNDS ARE SMALL and LOUD SOUNDS ARE UP – SOFT SOUNDS ARE DOWN. Sound concepts with a broader spectrum, e.g. articulation and timbre, were difficult to enact in coherent way. Although the results did not enter a formal requirements specification, the outlook was to use the results to inform the design of a learning system that can be used via whole-body interaction or via manipulating tangible objects.

Image schemas were also used to analyse and describe user interface metaphors (Kuhn & Frank, 1991). Zooming, for instance, was found to instantiate a NEAR-FAR image schema that mediates a PART-WHOLE image schema. Similarly, desktops and clipboards are instances of the SURFACE image schema; and folders and trashcans instantiate the CONTAINER image schema.

These studies show that it is possible to extract image schemas from user's utterances, user interfaces, the interaction with user interfaces, and even body movements. They do not show, however, that different analysts will come to the same conclusion as would be required when using image schemas as a metalanguage in design. Also, often the studies do not report how the insights of the analysis phase were used in the design phase.

4.1.2 Image Schemas in the Design and Evaluation Phases

In the third phase of the cycle, producing design solutions, image-schematic prescriptions could be instantiated in variants of a user interface concept that, after user evaluation, could be manifested in a concrete user interface design. Finally, in the fourth phase, image schemas can be used in the evaluation of user interfaces by comparing the image schemas instantiated in the user interface with the image schemas in the requirements specification. When they match, intuitive use of a system is more likely than when they do not match.

Image schemas and their metaphorical extensions have been proven effective in the design phase of prototypical applications. One application, SchemaSpace (Lund, 2003), is a collection of WWW bookmarks organised in a hierarchy (*Figure 4.2*). Semitransparent cones in an information landscape represent different categories of bookmarks, thus drawing on the metaphor CATEGORIES ARE CONTAINERS. The more bookmarks there are in one category, the taller the cone is (MORE IS UP). The relevance of single bookmarks in a category is conveyed by the metaphor IMPORTANT IS CENTRAL. Connections between cones (LINKS) indicate the relations between subcollections of bookmarks. Higher-level categories were located higher in the landscape (e.g., on a hill) and lower level categories were located lower in the landscape drawing on the metaphor ABSTRACT IS UP – CONCRETE IS DOWN. Finally, similar categories (cones) are located near each other and dissimilar items are located far from another (SIMILAR IS NEAR – DIFFERENT IS FAR).

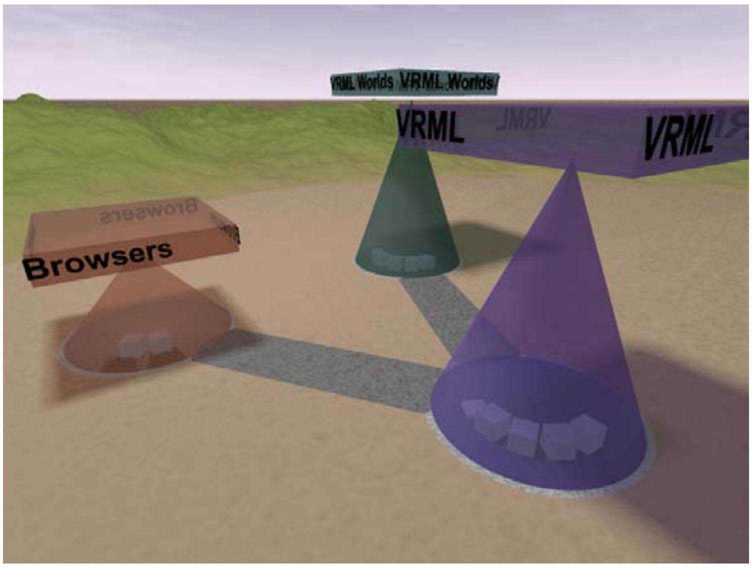


Figure 4.2. SchemaSpace, a personal information browser, the picture illustrates the image-schematic metaphors CATEGORIES ARE CONTAINERS, MORE IS UP and CONNECTEDNESS IS LINKAGE (Lund, 2003, p. 150)

The SchemaSpace prototype was evaluated with a number of users that solved information finding tasks and was compared with an information-equivalent hypertext prototype (Lund, 2003). The results show that the SchemaSpace prototype elicited significantly more comments from the users that contained the image schemas CENTRE-PERIPHERY, CONTAINER, LINK, NEAR-FAR, PART-WHOLE, PATH, and UP-DOWN than the hypertext prototype. The hypertext prototype, in contrast, elicited only more comments containing the image schema SURFACE. The author concludes that implementing metaphorical instantiations in user interfaces profoundly influences how users think about the interface. Unfortunately, no performance or satisfaction data are reported, so that it remains unclear what image schemas contribute to the different facets of intuitive use. A second study comparing ego-moving and time-moving instantiations of the metaphor THE PASSAGE OF TIME IS MOVEMENT THROUGH SPACE did not lead to clear results, which might be due to the concrete application used in this study (time-planning for mobile workers; Lund, 2003).

Based on the metaphor MUSIC IS BODY MOVEMENT, the Sound Maker system was built that enables users to interact with sound via moving their body (Antle, Droumeva, & Corness, 2008). Based on pilot studies and interviews with experts in music and movement, mappings between body movements and sounds were developed that were labelled either ‘metaphoric’ or ‘non-metaphoric’. These mappings were implemented in different versions of the Sound Maker system. After a free exploration phase, pairs of children (between 7 and 10 years old) listened to samples of music and were asked to re-enact these with the system. Results showed that children in the metaphoric condition could solve more tasks correctly and a little (although not significantly) faster than children in the non-metaphoric condition. There were also trends (albeit non-significant) in which children rated the metaphoric version easier to learn, more intuitive, and requiring less concentration to use. The main results on time and accuracy were later confirmed in a study with an adult sample (Antle, Corness, & Droumeva, 2009). Unfortunately, it can be debated whether the mappings labelled ‘non-metaphoric’ in these studies were not just alternative ‘metaphoric’ mappings. For example, LOUD SOUNDS ARE NEAR – SOFT SOUNDS ARE FAR can be a viable primary metaphor, because the loudness and the distance of objects are frequently related in our experience. This unclear status of the ‘metaphorical’ and ‘non-metaphorical’ versions of the systems renders the results difficult to interpret.

In another study, image-schematic metaphors were used to evaluate a software tool for playing, analysing, and learning about musical harmony (Wilkie, Holland, & Mulholland, 2009). The evaluation elicited conceptual metaphors and image schemas from the dialogues of experienced musicians discussing the harmonic progressions in a piece of music, e.g. HARMONIC PROGRESSION IS MOVEMENT ALONG A PATH and A CHORD IS A CONTAINER FOR NOTES. The authors then discuss where the software user interface supports the conceptual metaphors and where support could be improved.

Image schemas are not only useful for conveying functional information; they can also be used for conveying aesthetic information. The instantiation of the image schemas UP-DOWN, CONTAINER, and BALANCE was manipulated in jugs and alarm clocks (van Rompay, Hekkert, Saakes, & Russo, 2005). The image schema UP-DOWN, for instance, was instantiated by manipulating the height of the objects. Participants rated the their impression of the objects on nine dimensions, for example, secure–insecure, introvert–extravert, and constricting–liberating. The results showed that the image–schematic variations in product appearance influenced the ratings on the abstract dimensions. Unfortunately, the study was not based on specific metaphorical extensions derived from theory, so it provides no evidence for or against the validity of specific metaphors.

4.1.3 Conclusion

The evidence on using image schemas in user interface design shows that it is feasible to apply image schemas in the analysis of users' mental models and of user interfaces. Similarly, designing instantiations of image schemas into the user interface influences the perception of the product, how users solve and how they talk about their tasks. These results are encouraging for the design for intuitive use, but at the same time these studies are limited in three ways. First, according to our definition of intuitive use, it needs to be shown that applying image schemas and their metaphorical extensions leads to higher effectiveness and higher mental efficiency as well as to higher satisfaction. The studies of Antle et al. (2008, 2009) come closest to this goal, but the results are difficult to interpret. Second, in each of the studies of Lund (2003) and Raubal (1997) only one analyst conducted the extraction of image schemas. As Lund has pointed out, these image-schema extractions might therefore be very subjective. Different analysts might come to different solutions. It needs to be shown that there is reasonable inter-rater agreement in the application of image schemas. Third, none of these studies goes the full cycle of the human-centred design process (*Figure 4.1*). Image schemas were applied during context of use analysis (Bakker et al., 2009; Kuhn & Frank, 1991; Raubal, 1997), in deriving design solutions and for the evaluation of design solutions with users (Antle et al., 2008, 2009; Lund, 2003; Wilkie et al., 2009). But the transfer of image schemas into the requirements stage and from there to producing design solutions has not been studied so far. The transition from requirements to design solutions has been identified as a problem for design, because here the switch from analysis to synthesis must take place (cf. Dubberly, Evenson, & Robinson, 2008; Pahl, Beitz, Feldhusen, & Grothe, 2007). If the promise is for image schemas to act as a metalanguage for design, they must be able to cover all phases of the human-centred design cycle and be able to close the 'design gap' (Wood, 1998). Clearly, more evidence on this needs to be provided.

4.2 An Image-Schema Vocabulary

If image schemas are to be used as a metalanguage for designing user interfaces, there needs to be a relatively fixed vocabulary to start working with. Table 4.1 gives an overview of something like 'the standard inventory' of image schemas. The list draws from the inventories of image schemas given in the original philosophical work by Johnson (1987) and other image schemas derived from linguistic analyses (cf. Baldauf, 1997; Clausner & Croft, 1999; Hampe, 2005; Talmy, 2005), psychological studies (cf. Gibbs & Colston, 1995), and from the author's analyses of existing user interfaces (cf. chapter 8.2).

Table 4.1. *List of Image Schemas Used in the Empirical Studies*

Group	Image Schemas
BASIC	OBJECT, SUBSTANCE
SPACE	CENTER-PERIPHERY, CONTACT, FRONT-BACK, LEFT-RIGHT, LOCATION, NEAR-FAR, PATH, ROTATION, SCALE, UP-DOWN
CONTAINMENT	CONTAINER, CONTENT, FULL-EMPTY, IN-OUT, SURFACE
MULTIPLICITY	COLLECTION, COUNT-MASS, LINKAGE, MATCHING, MERGING, PART-WHOLE, SPLITTING
PROCESS	CYCLE, ITERATION, SUPERIMPOSITION
FORCE	ATTRACTION, BALANCE, BLOCKAGE, COMPULSION, COUNTERFORCE, DIVERSION, ENABLEMENT, MOMENTUM, RESISTANCE, RESTRAINT REMOVAL, SELF-MOTION
ATTRIBUTE	BIG-SMALL, BRIGHT-DARK, FAST-SLOW, HARD-SOFT, HEAVY-LIGHT, SMOOTH-ROUGH, STRAIGHT, STRONG-WEAK, WARM-COLD

One danger of such inventories of image schemas is that they are never complete. Johnson (1987), however, is confident that such lists capture the most important image schemas. Indeed, as the cross-cultural evidence shows, there seem to be universal sets of primitives that are close to the image schemas enlisted in Table 4.1 (Talmy, 2005; Wierzbicka, 1996).

It can be doubted, however, whether there are necessary and sufficient conditions to determine what qualifies as an image schema and what does not. As a consequence, it may appear that image schemas can only be specified by enumeration and accretion but not by prediction (cf. Clausner, 2005; Clausner & Croft, 1999). However, the definition and the list of image-schema characteristics (chapter 3.1.1) provide constraints to what can be regarded an image schema and what can not. For instance, because image schemas are very abstract representations, it is unlikely that concepts like DOG or CHAIR qualify, although these entities form recurring patterns of experience. As a second example, at a more abstract level, there is no image schema OBLIQUE, because it would be too weak to count as a dominant structure of experience. That becomes apparent when compared to the image schemas UP-DOWN and FRONT-BACK that are grounded in the experience of gravity (UP-DOWN) and the specifics of human anatomy determining the default directions of gaze and movement (FRONT-BACK). Studies on the naming of and memory for spatial relations (Hayward & Tarr, 1995) and on the representation of spatial dimensions in artwork (Appelle, 1972; Miller, 2007), for example, confirm that vertical and horizontal structures are more salient to human cognition than diagonal or oblique structures.

Another constraint on image schemas is that they tend to focus on physical properties of and relations between objects, so that, for example, recurring social experiences are not considered image schemas. Thus, there is, for example, no image schema FACE, although it would technically qualify. People are very responsive to faces from an early age on, faces are processed subconsciously, have internal structure, and are frequently used to describe and interpret objects (e.g. the cute or aggressive 'face' of a car). Similarly, recurring bodily actions such as seeing, jumping, or walking do not count as image schemas, as they are thought to be more foundational than image schemas (e.g. found on the procedural level in *Figure 3.4*; also cf. the concept of 'mimetic schemas', Zlatev, 2005).

The inductive, empirical nature of image schemas entails that the image schemas listed in Table 4.1 are not mutually exclusive. Although the grouping of image schemas suggests independence of one group from another, image schemas are better thought of as organised in a network instead of in a fixed hierarchy. The relationships between image schemas are manifold. Some image schemas depend on another, like RESTRAINT REMOVAL depends on BLOCKAGE. Others entail each other: The boundary of a CONTAINER can constitute a BLOCKAGE preventing the content leaking from the CONTAINER. Some image schemas depend on other image schemas in their instantiations. The image schema SCALE, for instance, is dependent on other image schemas like UP-DOWN, LEFT-RIGHT, or CENTRE-PERIPHERY. Instances of FORCE image schemas almost always include PATHS of motion (cf. Peña Cervel, 1999).

These mutual dependencies and interrelationships between image schemas can be a problem when applying them as the vocabulary of a metalanguage in user interface design. Any two designers might not agree on which image schema is instantiated in the user's mental model or the user interface, if they focus on different aspects of the situation (e.g. the cause and the effect of an event). Such disagreement can seriously hamper the application of image schemas as a design language. So the strength of image schemas, their experiential grounding (in contrast to a clear-cut vocabulary that is deductively derived from theory) could turn into a problem when applied in practise. One might argue, however, that the vocabulary in normal everyday language is not clear-cut either, and mutual comprehension is still possible.

It will be necessary to acknowledge the preliminary nature of such image-schema lists as given in Table 4.1 that require further tuning by empirical investigation.

Eventually, it is also an open empirical research task to determine whether the image-schema vocabulary in Table 4.1 can be used as a design language with reasonable agreement between speakers.

4.3 Research Questions and Scope of the Empirical Studies

Given the large potential of image schemas for user interface design and the limitations of previous studies, a potentially large number of research question emerges. The results of the previous discussion and the major demands for research are summarised as follows.

First, although many positive evaluations of conceptual metaphors are found in the cognitive science literature, there is almost no investigation of metaphorical extensions of image schemas in user interface design that applies measures of intuitive use. The results of such studies would help in judging the suitability of image schemas to meet the demand for abstract data.

Second, a large promise of image schemas is that they can be used as a metalanguage in user interface design. Previous studies have established that image schemas can be used in single phases of the human-centred design cycle. No study has shown, however, that image schemas can be applied continuously throughout the whole cycle, as is implied by their metalanguage character. Furthermore, little is known about the practicability of image schema use, expressed as the potential costs and benefits associated with using image schemas in practise.

Third, there is a lack of evidence on the stability of image-schema interpretations across different analysts. As image schemas can be highly interdependent, agreement between analysts could be low if they focus on different aspects of ambiguous stimuli. Therefore, the inter-rater agreement of image-schema analyses must be investigated as a prerequisite of applying image schemas as the vocabulary of a design language.

Fourth, possibly connected to the previous point, more research is needed on the features of artefacts that activate image schemas in the user's mind. Previous research in cognitive science found that whether an image-schematic interpretation of a spatial scene is achieved does not only depend on geometry and topology but also on other features like function, labelling, and so on.

Fifth, it is not clear whether image schemas and their metaphorical extensions are universal enough to contribute to user interfaces that can accommodate for large and heterogeneous user groups.

The studies in the following chapters are dedicated to provide data towards the first, the second and the third point. The first four experiments investigate the applicability of image-schematic metaphors to user interface design using measures of mental efficiency, effectiveness, and satisfaction. The fifth and the sixth study investigate the inter-rater agreement of people extracting image schemas from users' utterances, user interfaces, usage situations, and task descriptions – as would also be done when using image schemas in a human-centred design process. Finally, the seventh study investigates the practicability of image schemas in all phases of the redesign of an accountancy software package to estimate the costs and potential benefits of applying image schemas in a human-centred design process.

The results on inter-rater agreement already give an indication of how much variation to expect in the interpretation of image-schematic content. Nevertheless, it will be necessary to investigate when users interpret a specific instantiation of an image schema in a user interface as that image schema and when not (the fourth point). For example, what design features evoke a CONTAINER image schema in the users' mind and what differentiates them from features that evoke a SURFACE image schema? Although necessary and interesting, these questions are not addressed in the work presented here.

If the results on the applicability, reliability, and practicability of image schemas in user interface design are positive, the ground is prepared to follow on with the differential and intercultural aspects of image-schema use (the fifth point above). This will include research with user groups that are likely to differ in their kinds of prior knowledge, i.e. users with different cultural backgrounds, novices and experts, or users of different technology generations (old and young users). However, as the field has just begun to explore the usefulness of image-schema theory in user-centred design, investigating differential or intercultural aspects is beyond the scope of this work (but see Hurtienne, Langdon, & Clarkson, 2009; Hurtienne et al., 2010 for the discussion and application of image schemas in inclusive design).

5 Conveying Abstract Information with UP-DOWN and LEFT-RIGHT

Image schemas fulfil the preconditions of intuitive use (chapter 3.2) and they are involved in forming primary conceptual metaphors (chapter 3.1). Therefore, instantiating image-schematic metaphors in user interfaces should enhance the outcomes of intuitive use and make interaction with abstract data possible. A number of studies in cognitive science confirm the psychological reality of conceptual metaphors (chapter 3.3.1). However, systematic studies of image-schematic metaphors in user interfaces are rare (chapter 4.1).

This and the following chapters present experiments manipulating instances of image-schematic metaphors in simple user interfaces. The prediction is made that metaphor-congruent user interfaces are more intuitive to use than metaphor-incongruent user interfaces in terms of their effectiveness, mental efficiency, and user satisfaction.

5.1 Metaphorical Extensions of UP-DOWN and LEFT-RIGHT

Psychological evidence is available for a number of metaphorical extensions of the UP-DOWN image schema: HAPPY IS UP – SAD IS DOWN, POWER IS UP – BEING SUBJECT TO POWER IS DOWN, and GOOD IS UP – BAD IS DOWN (chapter 3.3.1). These studies showed that the vertical arrangement of stimuli influences response times when participants are processing the semantic content of abstract words (Meier & Robinson, 2004; Schubert, 2005). It was also shown that the processing of metaphorical and non-metaphorical abstract language interacts with UP-DOWN gesturing and arm movements (Casasanto & Lozano, 2006, 2007), and that the processing of non-verbal affective stimuli interacts with the remembered location of these stimuli on the UP-DOWN axis (Crawford et al., 2006). The results of these studies show that the metaphorical extensions of the UP-DOWN image schema were processed automatically and beneath consciousness.

The LEFT-RIGHT spatial axis is not as salient as the UP-DOWN (or FRONT-BACK) axis in human experience. The differences between LEFT and RIGHT are less pronounced than the differences between UP and DOWN (or FRONT and BACK). Animals, plants, and many artefacts are structured in a way that the UPPER and LOWER, the FRONT and the BACK parts differ from each other, while their LEFT and RIGHT SIDES are symmetrical. In our human bodies the functional differences between LEFT and RIGHT are not as profound as those between UP and DOWN (head and feet) or FRONT and BACK (default direction of visual perception and movement). Differences on the LEFT-RIGHT are more of a different degree than of a different kind. For example, the right hand is usually stronger or more dexterous than the left hand (but it still is a hand). The words *left* and *right* are learned the latest (Johnston, 1988) and are easily confused (Cienki, 1993), which is not the case for words denoting the UP-DOWN and FRONT-BACK axes.

Similarly, the metaphorical extensions of the LEFT-RIGHT image schema in language are not as strong and direct as the metaphorical extensions of the UP-DOWN image schema. The only ‘proper’ metaphorical extension of the LEFT-RIGHT image schema is the cultural convention of differentiating political parties into conservative *right-wing* parties and socially oriented *left-wing* parties. However, it has been noted that there is a loose association of the RIGHT with positive and the LEFT with negative qualities (Krzeszowski, 1993, 1997). Although there is talk like *He is my right hand* or *That was a left-handed compliment*, these metaphors less directly denote the target domain of quality as an UP-DOWN image schema can. Mostly these metaphorical extensions are restricted to one pole of the LEFT-RIGHT image schema. While one can talk about *the ups and downs of life*, there is no conventional talk about *him being my left hand* or a *right-handed compliment*.

It must be noted that there is a tendency to map physical UP to physical RIGHT and physical DOWN to physical LEFT (see Proctor & Vu, 2006). In abstract mappings, ‘spill-over’ effects of UP-DOWN metaphors to the LEFT-RIGHT axis can occur, at least for right-handers (Casasanto, 2009a). From the results of these studies it can be expected that metaphorical extensions of the UP-DOWN image schema will spill over to a RIGHT-LEFT orientation but that the RIGHT-LEFT effect may be weaker than the UP-DOWN effect.

There is a notable exception to this hypothesis. Although there is no linguistic evidence, a strong LEFT-RIGHT metaphor has been described in the target domain of quantity. This is the mapping MORE IS RIGHT – LESS IS LEFT, a well-known population stereotype (e.g. Proctor & Vu, 2006; Schmidtke & Rühmann, 1981). This population stereotype is not only present when asking users about their preferences but is also automatically triggered during online processing of numbers – a phenomenon known as the SNARC effect (Spatial Numerical Association of Response Codes). The SNARC effect refers to the observation that participants are faster to make a judgment about a number (e.g. whether it is odd or even) if the hand they use to respond is congruous with the size of the number in question. Responses with the left hand are quicker for smaller numbers and responses with the right hand are quicker for larger numbers (Gevers, Reynvoet, & Fias, 2003). For example, when instructed to respond to even numbers by pressing a button with their left hand, and to respond to odd numbers by pressing a button with their right hand, participants are quicker responding to *2* than to *8* and also quicker responding to *7* than to *3*. The SNARC effect points to the mental activation of a spatial number line with low numbers on the left and large numbers on the right. The SNARC effect was replicated in oculomotor studies (Fischer, Warlop, Hill, & Fias, 2004) and with neuropsychological patients suffering from hemispatial neglect (Loftus, Nicholls, Mattingley, & Bradshaw, 2008; Longo & Lourenco, 2007).

Two experiments are described in the following that attempt to verify the metaphorical extensions of the UP-DOWN and LEFT-RIGHT image schemas. In the first experiment, virtual push buttons are used to instantiate the image schemas. In the second experiment, virtual sliders are used as a different instantiation of these image schemas in the user interface.

5.2 Study 1: Virtual Push Buttons

The experiment includes four metaphorical extensions of the UP-DOWN image schema that are inspired by linguistic findings. The target domains of these metaphors are:

- Virtue, as in: *She is an upright citizen. That would be beneath me. He is high minded. That was a low-down thing to do*. (VIRTUE IS UP – DEPRAVITY IS DOWN)
- Quality, as in: *Things are looking up. This is a high-quality product. We hit a peak last year, but it's been downhill ever since*. (GOOD IS UP – BAD IS DOWN)
- Quantity, as in: *The number of books printed each year is going up. Sales rose last year. He is underage.* (MORE IS UP – LESS IS DOWN)
- Status, as in: *She'll rise to the top. He has little upward mobility. He's at the bottom of the social hierarchy.* (HIGH STATUS IS UP – LOW STATUS IS DOWN)

All of these target domains also play a role in the description of hotels – the domain of application chosen for this experiment. For instance, in a standard traveller guide (e.g. Harper et al., 2005) hotel descriptions include information on whether a hotel is a *luxury hotel* and whether it is *centrally located* (status). There is information about prices and the number of rooms in a hotel (quantity). Often, a summary evaluation is given, e.g. in saying that a hotel is *good value* (quality). Also, the quality of the staff is mentioned, e.g. *the hotel has a friendly owner who speaks excellent English* (virtue). And sometimes, the location of the facilities are described, e.g. *the restaurant is located at the top floor* (physical UP-DOWN).

In the experiment, participants were instructed to enter hotel evaluation data using a simple UP-DOWN button interface. The task was, for instance, to indicate whether the staff is friendly or unfriendly (*Figure 5.1*). In one condition, the user interface was congruent with the metaphor, e.g. VIRTUE IS UP, in a second condition it was incongruent with the metaphor. The hypothesis was that congruent button arrangements would be more intuitive to use than incongruent button arrangements.

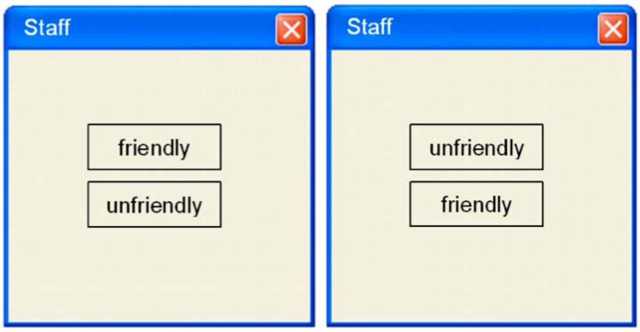


Figure 5.1. Examples of vertical button arrangements instantiating the metaphor VIRTUE IS UP (left: congruent with the metaphor, right: incongruent; translated from German).

Designing metaphor-congruent user interfaces should enhance all three criteria of intuitive use: effectiveness, mental efficiency, and user satisfaction. In terms of Rasmussen's model human information processor, the goal of evaluating the target domains of quantity, quality, virtue, and status is set by the experiment (module goal control in *Figure 2.1*) and activates these target domains in the dynamic world model. As these abstract target domains are assumed to be associated with the UP-DOWN image schema, the image schema becomes co-activated. In the metaphor-congruent user interfaces, the perception of the user interface matches the simulation in the dynamic world model, and the motor response can be triggered immediately. In a metaphor-incongruent user interface, perception does not match the simulation of the dynamic world model and the representations have to be re-organised before a response can occur. The process of reorganisation could take place as contention scheduling (chapter 2.3.2). Incongruent user interfaces therefore should lead to slower responses and they are lower in mental efficiency. Note that taking a time measure as an indicator of mental efficiency is appropriate here, because the experimental paradigm is simple enough to hold all other variables constant that could influence time readings as well (especially movements or other cognitive tasks that are not connected to the metaphors).

If the activation threshold of a metaphorical mapping is very low, responses can be directly triggered by activating the goal, without a prior update of the dynamic world model by perception. Hence, the error rate should be higher with metaphor-incongruent user interfaces (implying lower effectiveness) than with metaphor-congruent user interfaces.

Furthermore, if the user interface is metaphor congruent, it matches already existing mental associations. If users perceive reduced mental effort and a smoother interaction with the user interface, they should be more inclined to be satisfied with the interaction. Users should judge metaphor-congruent user interfaces to be more appropriate than user interfaces that are metaphor-incongruent.

In a control condition, the four target domains were combined with horizontal button arrangements instantiating a LEFT-RIGHT image schema as the source domain of the metaphor. In accordance with the literature discussed above, it is expected that metaphorical extensions of the LEFT-RIGHT image schema are less strong than the metaphorical extensions of the UP-DOWN image schema. Therefore no or only very weak effects are hypothesised for the LEFT-RIGHT image schema. However, because of the strong evidence for MORE IS RIGHT – LESS IS LEFT, the effect for the target domain of quantity should be stronger than for the other domains. The hypotheses are summarised in Table 5.1.

Table 5.1. *Hypotheses in Experiment 1*

Target Domain	Vertical Button Layout (Source domain UP-DOWN)	Horizontal Button Layout (Source domain LEFT-RIGHT)
Virtue	VIRTUE IS UP > VIRTUE IS DOWN	VIRTUE IS RIGHT ≥ VIRTUE IS LEFT
Quality	GOOD IS UP > GOOD IS DOWN	GOOD IS RIGHT ≥ GOOD IS LEFT
Quantity	MORE IS UP > MORE IS DOWN	MORE IS RIGHT > MORE IS LEFT
Status	HIGH STATUS IS UP > HIGH STATUS IS DOWN	HIGH STATUS IS RIGHT ≥ HIGH STATUS IS LEFT
Physical UP-DOWN	UP IS UP > UP IS DOWN	UP IS RIGHT ≥ UP IS LEFT

Note. The greater than (>) and greater than/equal (≥) signs refer to the expected degrees of intuitive use. Metaphorical extensions are abbreviated, e.g. GOOD IS UP – LESS IS DOWN is reduced to GOOD IS UP.

5.2.1 Method

Participants

Participants were 40 native German speakers (17 male, 23 female) recruited from the campus of Technische Universität Berlin. They were between 18 and 37 years old (*M* = 26.4, *SD* = 4.3). All participants were right-handed. The majority (55%) had stayed in one to five different hotels during the last 12 months. About a third (30%) did not stay in a hotel, 12.5% stayed in more than five hotels and 2.5% gave no details. Participation in the experiment took place in exchange for payment (eight Euros).

Procedure

At the beginning, participants completed a demographic questionnaire. They then received a written instruction in which they were told to enter data into a software program for the evaluation of 20 hotels (cf. appendix A.1). Each hotel was evaluated concerning ten different aspects, covering the five target domains of virtue, quality, quantity, status, and the control condition physical UP-DOWN. Each hotel was introduced with a screen displaying a picture of a city accompanied by the sentence *Please enter the data of the hotel in X* (*X* being the name of the city). When the participants were ready to proceed, they pressed the RETURN key.

For each hotel the procedure was like this: A priming statement, e.g. *The staff is friendly,* was presented for 2000 milliseconds. Then the priming statement was replaced with a dialogue box (cf. *Figure 5.1*) showing a vertical arrangement of buttons. Participants responded with the upper ('8') or lower key ('2') on a numerical keypad that mirrored the arrangement of buttons on the screen. After their response had been collected, the next trial started, i.e. the next priming statement appeared. Ten statements of each hotel were presented in random order.

After the first block of 20 hotels in the vertical button condition, a second block of 20 hotels followed in the horizontal button condition. These blocks were each preceded by ten practice trials (i.e. the data for one hotel) that did not enter the statistical analyses. Answer keys on the numerical keypad were arranged horizontally during the second block (keys '4' and '6'). Half of the participants received the vertical block of trials first and the horizontal block second. The other half received the two blocks in reversed order. Participants were instructed to respond as fast and accurately as possible.

After their task at the computer, participants were asked to fill in a questionnaire containing questions on how they judge the importance of different hotel characteristics and on judging the suitability of different button arrangements for entering hotel data (*Figure 5.2*, below). The experiment took about 45 minutes to complete.

Material

Stimuli. Each of the five target domains (quality, virtue, quantity, status, plus physical UP-DOWN) was represented by two types of evaluative statements about a given hotel (Table 5.2). The target domain of virtue, for instance, was represented by the statements *The staff is friendly* and *The staff is competent*. Other sentences expressed the opposite meaning, e.g. *The staff is unfriendly*.

The priming statements of the target domain *virtue* were inspired by a list of virtues provided by Laug (2003). The statements for the target domain *status* were derived from a pre-test in which 12 participants (different from the participants in the main study) indicated what determines the social or societal status of a hotel. The most frequent answers regarded the location of the hotel (city centre versus periphery) and features that could be summarised under the categories 'luxury' and 'standard'. The target domain of quality was operationalised by good-bad judgments, and the target domain of quantity involved numbers. The specifics of each statement (whether it was positively or negatively phrased) and the corresponding button labelling (whether it was congruent or incongruent with the metaphor) were randomly assigned to the hotels.

Table 5.2. *Target Domains and Stimuli Used in Experiment 1*

Target domain	Priming statements	Button labels
Virtue	The staff is friendly / unfriendly. The staff is competent / incompetent.	friendly, unfriendly competent, incompetent
Quality	The breakfast buffet is good / bad. The rail connections are good / bad.	good, bad good, bad
Quantity	The hotel is booked by 90% / 70%. The parking garage has 100 / 30 spaces.	90%, 70% 100, 30
Status	The hotel is in the city centre / suburbs. The hotel is a luxury / standard hotel.	city centre, suburbs luxury, standard
Physical UP-DOWN	The hotel bar is up / down. The meeting rooms are up / down.	up, down up, down

Note. Stimuli are translated from German.

Source domains were reflected in the two possible button layouts – vertical (UP-DOWN) and horizontal (LEFT-RIGHT). Mappings between source and target domain could be either congruent or incongruent. Congruent mappings were instances of the respective metaphor (e.g. GOOD IS UP – BAD IS DOWN) or referred to the spill-over effect in the horizontal case (e.g. GOOD IS RIGHT – BAD IS LEFT).

Apparatus. Errors and response times were computed from the logfile of the computerised experiment. A notebook computer, Acer 4501LCi (Intel Centrino M715, 1.5-GHz, 15 inch display size, 1024x768 resolution), was used. Stimuli were delivered and responses were collected using the software package E-Prime 1.1. The centres of the two buttons were 12 mm (vertical buttons) and 37mm (horizontal buttons) apart. The viewing distance was 60 cm. For the labels, a 12 pt Arial font was used.

Questionnaire. Suitability judgments were collected with a questionnaire that consisted of ten pages. At the top of each page the two sentences of each priming statement were printed, e.g. *The staff is friendly* and *The staff is unfriendly* (see Table 5.2). For half of the participants the order of these statements was reversed. Then, the four variations of button layouts were presented reflecting the source domain x mapping variations (*Figure 5.2*). Participants indicated for each button layout how suitable it was for displaying the specific type of data indicated by the priming statement on top of the page. The five-point answering scale was coded from -2 (very ill suited) to +2 (very well suited).

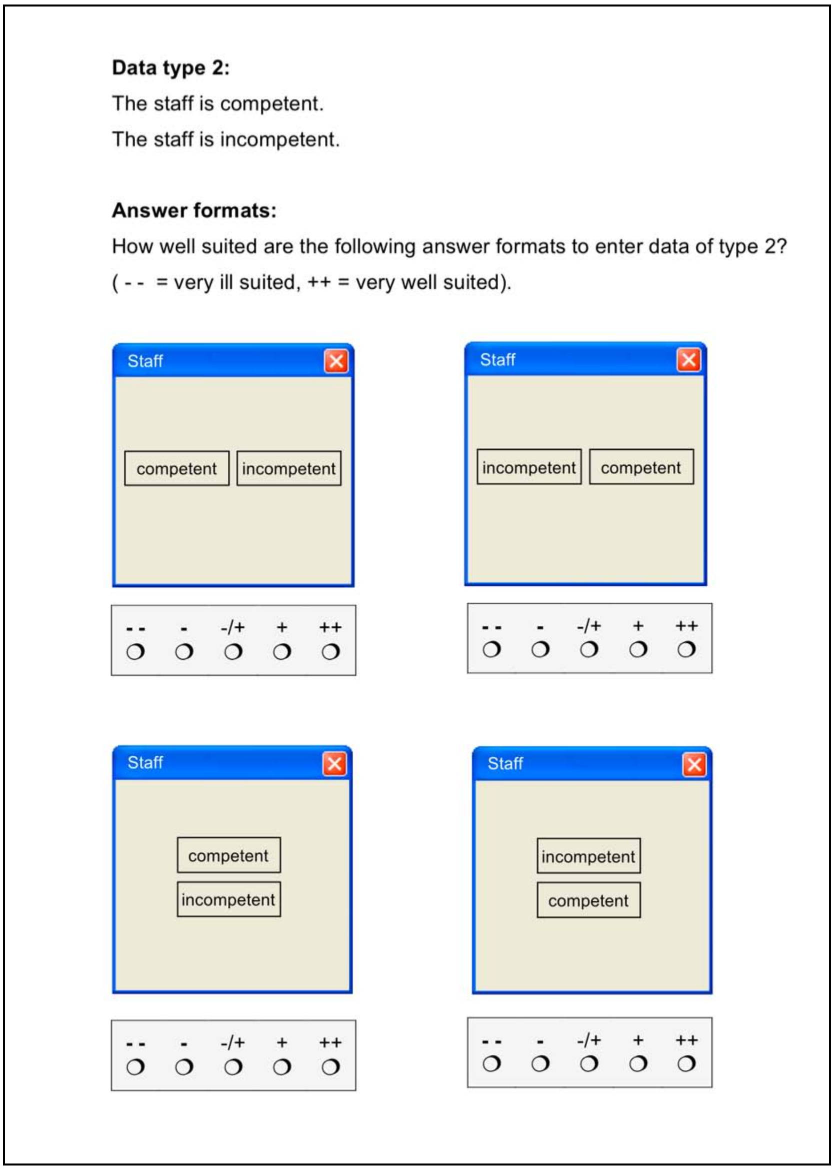


Figure 5.2. Sample page of the questionnaire for collecting suitability judgments in experiment 1 (translated from German)

Auxiliary data was collected with a demographic questionnaire asking for sex, age, handedness, highest level of education, job title, native tongue, and foreign languages, including the level of proficiency in speaking these languages (i.e. years spoken). An additional question asked for the number of different hotels stayed in during the last 12 months. On a separate questionnaire page participants were asked *How important are the following criteria for you in evaluating a hotel?* Then a list with the target domain statements was presented. Each of these statements was rated using a five-point scale ranging from 1 – not at all important to 5 – very important (for detailed results see appendix A.1.4).

Experimental Design

A within-subjects design was used with three independent variables: target domain (virtue, quality, quantity, status, and physical UP-DOWN), source domain (UP-DOWN and LEFT-RIGHT), and mapping (congruent and incongruent). According to the hypotheses generated above, the dependent variable is degree of intuitive use. In chapter 2 it was suggested that intuitive use influences effectiveness, mental efficiency, and satisfaction. Error rates are used as an indicator of effectiveness. Response times are used as indicators for mental efficiency, because in this experiment with all other conditions held constant, the differences in response times indicated differences in mental effort. Suitability judgments are used as indicators of satisfaction. Error rates and response times are inversely proportional to intuitive use, i.e. the more errors participants make and the longer they take to respond the less intuitive the design is. Suitability judgments are proportional to intuitive use, i.e. the more suitable a design is rated, the more intuitive to use it is.

Data analysis was conducted using E-Prime 1.1, SPSS 15.0, and an effect size calculation tool by Jacobs (2005). An effect size *d* of .20 would be regarded as small, of .50 as medium and of .80 as high (Bortz & Döring, 2002). To filter out non-task related responses, response times were truncated when they exceeded three standard deviations above the overall mean. Only response times from correct trials entered the analysis. Outliers were determined using the boxplot criterion (cf. Howell, 1997). A total of 1.9% of the data across all dependent variables were outliers that were replaced by mean values at the subject level. The GLM (General Linear Models) repeated measures procedure was used as an overall statistical test. When the Mauchly test of sphericity was significant, the degrees of freedom were corrected after Greenhouse-Geisser. For single comparisons, *t*-tests for repeated measurements were used and Bonferroni-Holm corrections were applied to achieve a family-wise error rate of $\alpha = .05$ (Howell, 1997). The error bars in all figures represent standard errors of the mean.

5.2.2 Results

Metaphor-congruent button labels are expected to lead to less errors, faster response times, and higher subjective suitability ratings than incongruent button labels. These effects are expected to be smaller or not apparent in the condition with the horizontal button layout. An exception is the target domain of quantity for which stronger effects are expected according to the mapping MORE IS RIGHT – LESS IS LEFT.

The following sections present the results in detail. The low difficulty of the task led to a ceiling effect regarding error rates. Overall accuracy was as high as 98.1%. The majority of 72% of the participants made no errors at all (averaged across conditions). As this ceiling effect compromises the meaningfulness of statistical results, no further error-rate statistics are reported. For all other data tables see appendix A.1.

Vertical Button Arrangements

Between mapping conditions, the overall effect of response times as the indicator for mental efficiency was as expected. Congruent mappings had significantly lower response times than incongruent mappings, $F(1,39) = 33.16$, $p < .001$. Further, there were significant effects of target domain, $F(4,156) = 82.42$, $p < .001$, and of the interaction between target domain and mapping, $F(4,156) = 9.50$, $p < .001$.

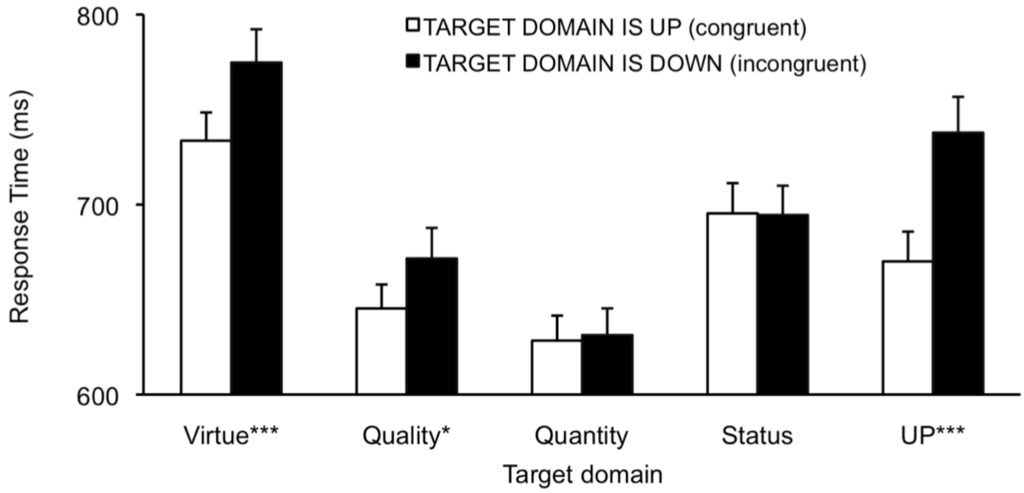


Figure 5.3. Experiment 1, vertical button arrangements: Response times for different target domains and mapping conditions. * p < .05, *** p < .001 (Paired t-tests, Bonferroni-Holm corrected)

Single comparisons show that response times in metaphor-congruent mappings were faster than in incongruent mappings in the target domains virtue, $t(39) = -4.16$, $p < .001$, $d = .40$; quality, $t(39) = -2.58$, $p < .05$, $d = .29$; and physical UP-DOWN, $t(39) = -5.61$, $p < .001$, $d = .62$. There were no significant differences in response times for the target domains quantity, $t(39) < 1$, and status, $t(39) < 1$ (*Figure 5.3*).

As expected, suitability judgments as the indicator of satisfaction were higher for congruent mappings and lower for incongruent mappings, $F(1,39) = 181.51$, $p < .001$. Further, there were significant effects of target domain, $F(4,156) = 4.21$, $p < .01$, and of the interaction between target domain and mapping, $F(2.40,93.66) = 38.35$, $p < .001$.

Single comparisons revealed that metaphor-congruent mappings were judged as more suitable than the metaphor-incongruent mappings in all target domains: virtue, $t(39) = 10.84$, $p < .001$, $d = 2.45$; quality $t(39) = 14.71$, $p < .001$, $d = 3.17$;

quantity, $t(39) = 1.86$, $p < .05$, $d = .51$; status, $t(39) = 3.41$, $p < .01$, $d = .68$; and physical UP-DOWN, $t(39) = 17.05$, $p < .001$, $d = 4.07$ (*Figure 5.4*).

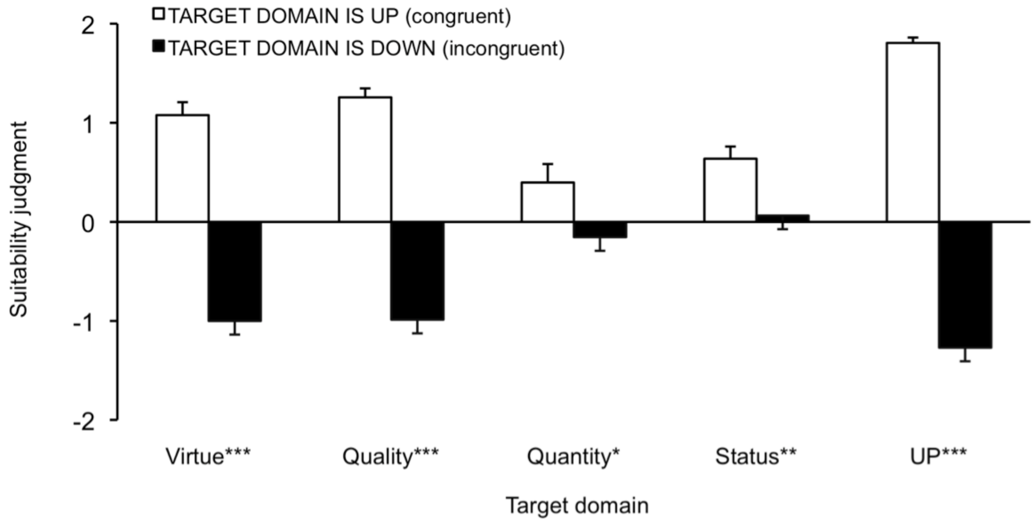


Figure 5.4. Experiment 1, vertical button arrangements: Suitability judgments for different target domains and mapping conditions. * p < .05, ** p < .01, *** p < .001 (Paired t-tests, Bonferroni-Holm corrected)

Horizontal Button Arrangements

Response times (mental efficiency) were not different between metaphor-congruent and metaphor-incongruent mappings $F(1,39) < 1$, n.s. Further, there were significant effects of target domain, $F(3.22,127.67) = 54.30$, $p < .001$, and of the interaction between target domain and mapping, $F(4,156) = 5.76$, $p < .001$.

Single comparisons show significant differences between metaphor-congruent and metaphor-incongruent mappings for quality, $t(39) = -2.55$, $p < .05$, $d = .31$, and physical UP-DOWN, $t(39) = 2.94$, $p < .05$, $d = .38$. Note that the effect for quality is in the opposite direction than predicted. Participants responded faster to a GOOD IS LEFT mapping ($M = 626.87$, $SD = 73.50$) than to a GOOD IS RIGHT mapping ($M = 652.36$, $SD = 88.92$). The effects for physical UP-DOWN were in the hypothesised direction (the UP IS RIGHT mapping was faster than the UP IS LEFT mapping). There were no significant differences in response times for the target domains virtue, quantity, and status, all $t(39) < 1$ (*Figure 5.5*).

Regarding suitability judgments (satisfaction), metaphor-congruent mappings were not different from metaphor-incongruent mappings in general, $F(1,39) < 1$, n.s. However, there were mapping differences at the level of single target domains that pointed in different directions causing the interaction effect between target domain and mapping to be significant, $F(2.38,92.66) = 23.89$, $p < .001$. There was also a significant effect for target domain, $F(3.25,126.59) = 3.14$, $p < .05$.

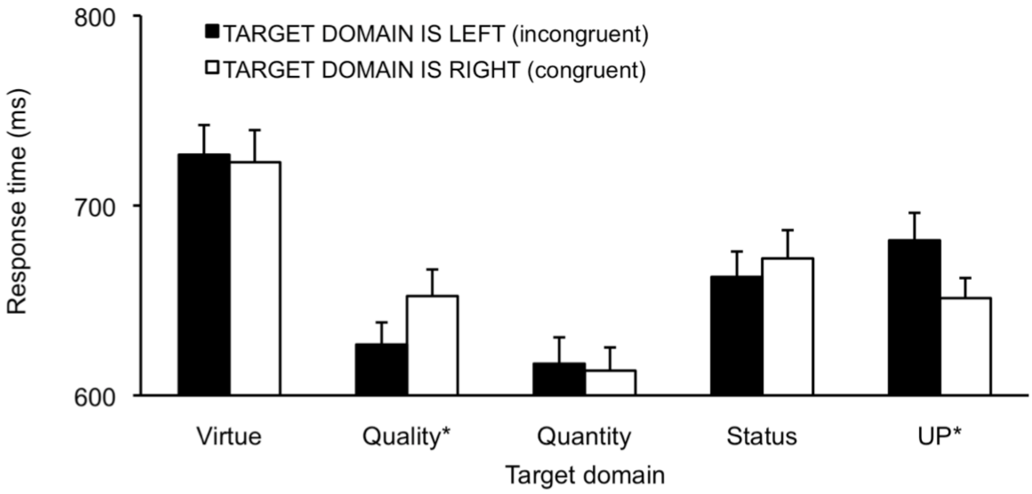


Figure 5.5. Experiment 1, horizontal button arrangements: Response times for different target domains and mapping conditions. * p < .05 (Paired t-tests, Bonferroni-Holm corrected)

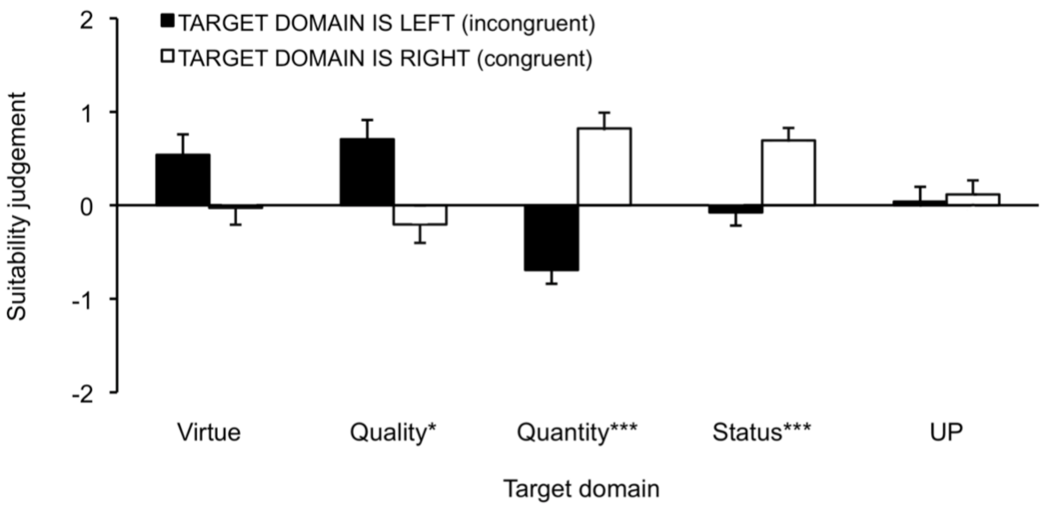


Figure 5.6. Experiment 1, horizontal button arrangements: Suitability judgments for different target domains and mapping conditions. * p < .05, *** p < .001 (Paired t-tests, Bonferroni-Holm corrected)

Single comparisons revealed that the hypotheses were confirmed for the target domains of virtue, quantity, status, and physical UP-DOWN, but not for quality. More specifically, congruent mappings were preferred over incongruent mappings for quantity, $t(39) = -5.97$, $p < .001$, $d = 1.50$, and status, $t(39) = -4.37$, $p < .001$, $d = .89$. Also as expected, no differences of suitability judgments were found in the mapping conditions of the target domains virtue, $t(39) = 1.57$, n.s., and physical UP-DOWN, $t(39) = -0.38$, n.s. Quite remarkably, suitability judgments for the target domain of quality followed a reversed pattern, identical to the one found in response times. Incongruent mappings (GOOD IS LEFT – BAD IS RIGHT) were preferred over congruent mappings (GOOD IS RIGHT – BAD IS LEFT), $t(39) = 2.57$, $p < .05$, $d = .72$ (*Figure 5.6*).

5.2.3 Discussion

The results of the first experiment show that image-schematic metaphors do make a difference in interacting with technology. First the overall results and then the results of the single comparisons are discussed here. Following the results of the overall test, it is concluded that the predictions of image-schema theory were confirmed for the UP-DOWN image schema (Table 5.3). Theory-congruent user interfaces were more intuitive to use than theory-incongruent user interfaces. They led to a higher mental efficiency and satisfaction of users. To derive meaningful effectiveness data, though, the task was too simple and almost no errors were made.

In the horizontal condition, no differences were found between the metaphor-congruent and the metaphor-incongruent mappings. The hypothesis expecting no or only weak effects with LEFT-RIGHT mappings was supported by the data on the general level (Table 5.4).

At a greater level of detail the results are more differentiated. The metaphorical extensions of the UP-DOWN image schema are discussed first (Table 5.3). The response time data confirmed the hypotheses for two metaphorical target domains, virtue and quality, as well as for the physical UP-DOWN condition, but not for quantity and status metaphors. The data on suitability judgments support the hypotheses for each target domain.

There might be two reasons why the target domain of quantity failed to support the hypothesis in the response time data. First, the hotel evaluation task could have put too much a focus on evaluating valence. The three target domains virtue, quality, and status emphasise thinking about positive or negative hotel characteristics. If intermediate trials presented numbers, participants may have also seen them in the light of evaluating valences. Thus, instead of invoking the target domain of quantity, the numbers on the buttons may have led to an attempt to interpret them in terms of valence.

Table 5.3. *Summary of Predictions and Results for the Vertical Button Arrangements*

	Prediction	Results	
Target domain	Intuitive Use	Mental Efficiency	Satisfaction
Overall	TARGET DOMAIN IS UP > TARGET DOMAIN IS DOWN	TARGET DOMAIN IS UP > TARGET DOMAIN IS DOWN ☑	TARGET DOMAIN IS UP > TARGET DOMAIN IS DOWN ☑
Virtue	VIRTUE IS UP > VIRTUE IS DOWN	VIRTUE IS UP > VIRTUE IS DOWN ☑	VIRTUE IS UP > VIRTUE IS DOWN ☑
Quality	GOOD IS UP > GOOD IS DOWN	GOOD IS UP > GOOD IS DOWN ☑	GOOD IS UP > GOOD IS DOWN ☑
Quantity	MORE IS UP > MORE IS DOWN	MORE IS UP = MORE IS DOWN ☒	MORE IS UP > MORE IS DOWN ☑
Status	STATUS IS UP > STATUS IS DOWN	STATUS IS UP = STATUS IS DOWN ☒	STATUS IS UP > STATUS IS DOWN ☑
Physical UP-DOWN	UP IS UP > UP IS DOWN	UP IS UP > UP IS DOWN ☑	UP IS UP > UP IS DOWN ☑

Note. Results are marked with ☑ if the hypothesis was confirmed, and ☒ if not confirmed.

In addition, the domain of quantity had low salience in the hotel evaluation context. The two number items referred to 'number of parking spaces' and 'percentage of bookings'. However, assigning valence to these numbers is quite difficult. For example, is a high percentage of bookings good (characterizing a popular hotel) or bad (no rooms available when visiting spontaneously)? In the questionnaire, participants judged the number of parking spaces and the percentage of bookings as not important for evaluating hotel quality, $M = 1.68$ and $M = 1.90$, respectively (on a scale from 1 - not at all important to 5 - very important; see appendix A.1.4).

Another interpretation is that the presentation of buttons in the user interface evoked CONTAINER image schemas in the user's mind. The numbers on the buttons were then interpreted as discrete entities representing different qualities, rather than as information that varies along a quantitative SCALE.

Similar reasoning can be applied to explaining why the target domain of status failed to support the hypothesis in the response time data. Although the participants judged the status of a hotel as fairly important ($M = 3.18$ and $M = 4.08$ for both items, appendix A.1.4), there is no data on the specific preferences of users. As the participants were mainly students, to them the high or low status of a hotel had a different salience than intended. As some participants remarked, their pragmatic decision is not to stay in a high status hotel because students usually travel on tight budgets. Both target domains – valence and status – might have worked against each other in the case of the status expressions. Hence, no effect in the performance data was found.In summary, using a hotel-evaluation context probably led to domain-dependent results. The strong focus on qualitative evaluation and the differing salience of specific hotel characteristics has influenced the processing of quantity and status stimuli. This is in line with the findings discussed in chapter 3 that emphasise the task-dependency of metaphorical effects on response time data (e.g. Schubert, 2005). In the target domain of quantity, using buttons instantiating CONTAINER image schemas might have reinforced these effects. Both issues are addressed in study 2: the context-dependency is removed and sliders are used instead of buttons.

What about LEFT-RIGHT as a source domain (Table 5.4)? It was hypothesised that there are spill-over effects from UP-DOWN to RIGHT-LEFT. These spill-over effects were not found in the overall test, but in some of the results for single target domains: status (suitability judgments) and physical UP-DOWN (response times). Because of its prominence in non-verbal communication, the MORE IS RIGHT – LESS IS LEFT metaphor was assumed to elicit stronger effects. Strong effects could be found in the satisfaction data, but not in the response time data. Here again, either using buttons as representations or the overall context of evaluation might have hindered the activation of the target domain of quantity.

Most surprising are the results for the target domain of quality. Here, the prediction GOOD IS RIGHT – BAD IS LEFT was not confirmed. Rather, in response times and subjective ratings, a GOOD IS LEFT – BAD IS RIGHT mapping seems more intuitive to participants.

Are alternative explanations of the results likely? First, it might be possible that the effects do not point to an effect of conceptual metaphor, but are an effect of word length or the frequency with which words are used in everyday language. Metaphor congruency effects thus could just be a matter of reading time. When the short or more frequent word comes first (i.e. *gut*), participants will be quicker to respond. Word lengths (measured as number of letters) did not differ, Wilcoxon test, $Z = -1.85$, n.s. Word frequencies were determined using the database *Wortschatz Universität Leipzig* (http://wortschatz.uni-leipzig.de, for details see appendix A.1.5). The word frequencies differed, $Z = -2.12$, $p < .05$. In the congruent labelling condition, more frequent words tended to appear at the top and less frequent words at the bottom.

However, if the results obtained are an effect of word frequency, the pattern of results should be the same in the horizontal and vertical button conditions. As the patterns of response-time results differ between these conditions, they do not support such a conclusion well. Only the results of the target domain of quality would support such an alternative explanation. But it is not clear then, why similar target domains, e.g. virtue, do not show similar results – as would be predicted if the word frequency hypothesis was correct.

Table 5.4. *Summary of Predictions and Results for the Horizontal Button Arrangements*

	Prediction	Results	
Target domain	Intuitive Use	Mental Efficiency	Satisfaction
Overall	TARGET DOMAIN IS RIGHT ≥ TARGET DOMAIN IS LEFT	TARGET DOMAIN IS RIGHT = TARGET DOMAIN IS LEFT ☑	TARGET DOMAIN IS RIGHT = TARGET DOMAIN IS LEFT ☑
Virtue	VIRTUE IS RIGHT ≥ VIRTUE IS LEFT	VIRTUE IS RIGHT = VIRTUE IS LEFT ☑	VIRTUE IS RIGHT = VIRTUE IS LEFT ☑
Quality	GOOD IS RIGHT ≥ GOOD IS LEFT	GOOD IS RIGHT < GOOD IS LEFT ☒	GOOD IS RIGHT < GOOD IS LEFT ☒
Quantity	MORE IS RIGHT > MORE IS LEFT	MORE IS RIGHT = MORE IS LEFT ☒	MORE IS RIGHT > MORE IS LEFT ☑
Status	STATUS IS RIGHT ≥ STATUS IS LEFT	STATUS IS RIGHT = STATUS IS LEFT ☑	STATUS IS RIGHT > STATUS IS LEFT ☑
Physical UP-DOWN	UP IS RIGHT ≥ UP IS LEFT	UP IS RIGHT > UP IS LEFT ☑	UP IS RIGHT = UP IS LEFT ☑

Note. Results are marked with ☑ if the hypothesis was confirmed, and ☒ if not confirmed.

Second, there might be a general ability of people to process positively evaluated stimuli quicker than negatively evaluated stimuli. This should result in shorter response times for positive words than for negative words in the first position. Again, such an effect should be equal for the vertical and the horizontal condition. As the response-time results show, the pattern of differences between mapping conditions in the vertical condition is not reflected in the horizontal condition. It seems that both alternative explanations are not supported by the data and the results are better explained by the conceptual metaphor hypothesis.

In conclusion, some predictions made for the metaphorical extensions of the image schemas UP-DOWN and LEFT-RIGHT can be assumed to be valid in a user interface design context. Alternative explanations are not very likely. Effect sizes (*d*) vary from very strong to medium for satisfaction data and from medium to small for mental efficiency. Some of the more detailed results for the target domains of quality, status, and quantity deviate from the overall predictions. A replication of the study using sliders instead of buttons and a context-neutral task might be able to clarify open issues found mainly in the response time data.

5.3 Study 2: Virtual Sliders

As some of the results in the previous study were difficult to interpret, a second experiment was conducted as an attempt to replicate the key issues. An unexpected outcome of the first experiment was that there were no response time differences for the different mapping conditions in the target domain of quantity. In response times, the mappings MORE IS UP / RIGHT – LESS IS DOWN / LEFT could not be confirmed. It was hypothesised that this is an effect of the hotel evaluation task that shifted attention to valence rather than quantity. A second hypothesis was that the presentation of buttons instantiated CONTAINER image schemas that distracted from activating the target domain of quantity.

In the second experiment, the evaluation context was removed and participants reacted to simple requests to change the value of a variable to be *more* or *less* (target domain quantity), *better* or *worse* (target domain valence), *higher* or *lower* (physical UP-DOWN control condition). The target domain of valence was introduced to reflect the coherence in mapping of many metaphorical extensions of the UP-DOWN image schema. For instance, the target domains of quality, health, life, virtue, happiness, coherently map to UP-DOWN described by the metaphor POSITIVE VALENCE IS UP – NEGATIVE VALENCE IS DOWN (Krzeszowski, 1993, 1997).

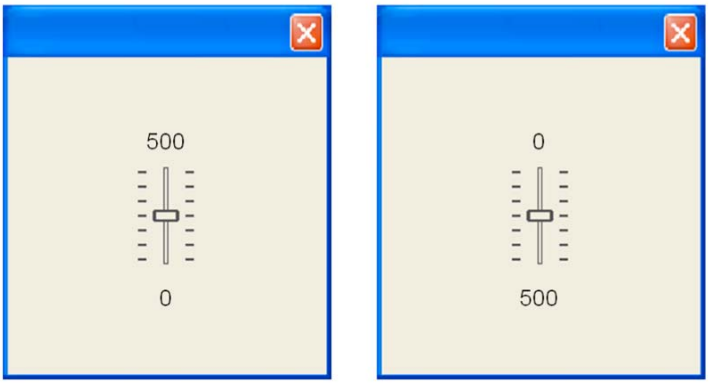


Figure 5.7. Examples of vertical slider labels instantiating the metaphor MORE IS UP (left: congruent with the metaphor, right: incongruent).

To emphasise the quantitative aspect in the user interface design, the virtual buttons were replaced by virtual sliders (*Figure 5.7*). Sliders were chosen as instances of the SCALE image schema. It was hypothesised that a SCALE image schema supports the activation of the target domain of quantity better than the CONTAINER image schema instantiated by the buttons in the previous experiment. Again, the same target domains were used with horizontal sliders instantiating the LEFT-RIGHT image schema. The hypotheses are the same as in experiment 1 (Table 5.5).

Table 5.5: *Hypotheses in Experiment 2*

Target Domain	Vertical Slider Layout (Source domain UP-DOWN)	Horizontal Slider Layout (Source domain LEFT-RIGHT)
Valence	POSITIVE VALENCE IS UP > POSITIVE VALENCE IS DOWN	POSITIVE VALENCE IS RIGHT ≥ POSITIVE VALENCE IS LEFT
Quantity	MORE IS UP > MORE IS DOWN	MORE IS RIGHT > MORE IS LEFT
Physical UP-DOWN	UP IS UP > UP IS DOWN	UP IS RIGHT ≥ UP IS LEFT

Note. The greater than (>) and greater than/equal (≥) signs refer to the expected degree of intuitive use.

5.3.1 Method

Participants

Participants were 40 native German speakers (21 male, 19 female) recruited from the campus of Technische Universität Berlin. They were between 19 and 47 years old ($M = 25.4$, $SD = 5.8$), and all were right-handed. Participation took place in exchange for payment (eight Euros).

Procedure

First, participants received a written instruction that informed them about the experimental procedure in general (see appendix A.2). The first part of the experiment was a short training at the computer consisting of 32 trials (16 trials each in the horizontal and vertical slider conditions) to get used to the procedure of the experiment. Then participants filled in a demographic questionnaire. After this the actual experiment started.

At the beginning of each trial participants were reminded to press the centre key on a numerical keypad (corresponding to the key *5*). While they held the centre key pressed, a single word priming a particular target domain was presented in the centre of the screen, e.g. *better*, *less*, *higher*. After 1300 milliseconds the word was replaced by a dialogue box with a vertical slider (as in *Figure 5.7*). Participants decided on the correct answer. For instance, if they were primed with the word *less* and the slider was labelled congruent with the metaphor MORE IS UP, the correct response would be to move the slider downwards. To respond, participants released the centre key and pressed the appropriate key above or below the centre key on the keypad. As soon as the centre key was released, both slider labels were each masked by a string of *XXXXXXXX*. This was done to prevent participants referring to slider labels after releasing the centre key. After the response had been collected, the next trial started. Trials were presented in random order.

After every 30 trials a short break was offered to the participant that could be ended by pressing the RETURN key. After a block of 120 trials with vertical sliders, the second block of 120 trials presenting horizontal sliders started. In the horizontal condition, responses were made with keys to the left and right of the centre key on the numerical keypad. Half of the participants received the block with vertical sliders first, the other half received the block with horizontal sliders first. They received a written instruction announcing the change in experimental stimuli before each block. Participants were instructed to respond as fast and accurately as possible.

After the task at the computer, participants filled in a questionnaire in which they were asked to judge the suitability of different slider arrangements for entering data (see below, *Figure 5.8*). Then they filled in an additional questionnaire on the aesthetic appearance of slider labels. The experiment took about 45 minutes to complete.

Material

Stimuli. The target domain valence was always primed by the two statements *better* and *worse* (in German). For the labels of the sliders, pairs of adjectives were used that were tested for their valence in a preparatory study. Based on German word standards (Ostendorf, 1994; Schwibbe, Räder, Schwibbe, Borchardt, & Geiken-Pophankem, 1994) a number of 40 adjective pairs were selected. Of these, 28 pairs were rated using a paper-and-pencil questionnaire. The remaining 12 pairs were rated using an online questionnaire. Participants (not the same as in the main experiment) rated each adjective on how much it is related to the domains of quality and quantity (1 – not at all to 7 – very strongly). In addition they evaluated the adjectives on the dimensions unpleasant-pleasant and bad-good. Data from 45 participants in the online study and 54 participants in the paper-and-pencil study entered the analyses. From the results, adjectives were selected that were strongly related to quality and not or only weakly related to quantity. Furthermore, of each pair of adjectives one adjective had to be rated significantly good and pleasant, the other significantly bad and unpleasant. Of the remaining 18 adjective pairs, 12 pairs were included in the experiment. Examples include the German equivalents of *inventive-boring, satisfied-dissatisfied,* and *practical-impractical* (see appendix A.2.2).

The target domain of quantity was always primed by the two statements *more* and *less* (in German). For the labels of the sliders 12 pairs of numbers were randomly determined that differed in length and roundedness (appendix A.2.2). In order to enable participants to use the same strategy that could be applied to adjective labels, the pairs of number labels always included a *0*. Thus, participants reading one label could infer the value of the other that could be zero or non-zero. Example label pairs include *5-0, 289-0,* and *8000-0* (see appendix A.2.2).

The target domain of physical UP-DOWN was always primed by the two statements *higher* and *lower* (in German). For the slider labels six pairs of adjectives describing physical UP-DOWN relations were determined using MS Word's thesaurus function. Examples include the German equivalents of *up-down, above-below*, and *on top-underneath* (see appendix A.2.2).

The image-schematic source domains were instantiated in the two possible slider layouts – vertical (UP-DOWN) and horizontal (LEFT-RIGHT). Mappings between source and target domain could be either congruent or incongruent. Congruent mappings were compatible with the respective metaphor (e.g. POSITIVE VALENCE IS UP – NEGATIVE VALENCE IS DOWN) or with the spill-over effect in the horizontal case (e.g. POSITIVE VALENCE IS RIGHT – NEGATIVE VALENCE IS LEFT).

Apparatus. The same equipment as in experiment 1 was used. The centres of the two slider labels were 37 mm (vertical) and 47 mm (horizontal) apart. The viewing distance was 60 cm. For the labels, a 12 pt Arial font was used.

Questionnaires. Suitability judgments were collected with a questionnaire (*Figure 5.8*). At the top of each page a different question for each target domain was printed, e.g. for the domain of quantity *How well suited, do you think, are the following sliders to make adjustments in the directions "more" or "less"?* (Half of the participants received a version of the questionnaire with the two target words in reversed order, i.e. *…"less" or "more"?*). Four variations of slider layouts were presented reflecting the 2x2 source domain x mapping variations (*Figure 5.8*). Participants indicated for each slider layout how suitable they found it for displaying the specific type of data. The five-point answering scale was coded from -2 (very ill suited) to +2 (very well suited). Each participant received a different randomised order of the questionnaire pages.

The same demographics questionnaire as in the first study was used. Two additional questions asked for the number of years of computer experience and the frequency of using computers in hours per week (not reported here). An additional questionnaire was used to rate the aesthetic appearance of slider labels. The results of this questionnaire failed to deliver a possible alternative explanation of the findings of this experiments and thus are not reported here. For further details and results on this questionnaire see Loest (2007).

Experimental Design

A within-subjects design was used with three independent variables: target domain (valence, quantity, and physical UP-DOWN), source domain (UP-DOWN and LEFT-RIGHT), and mapping (congruent and incongruent).

Again, the global dependent variable was the degree of intuitive use. The same measures were taken as in experiment 1 – error rates as an indicator of effectiveness, response times as indicators of mental efficiency, and suitability judgments as indicators of satisfaction. Errors and response times were taken from the computerised experiment. Response times were measured from the onset of the dialogue box to the release of the centre button. This made sure that the time for mental processes was measured independently from the time needed to actually navigate the finger to and press the correct response button. Data analysis was conducted using the same tools and procedures as in study 1. No outliers were removed.

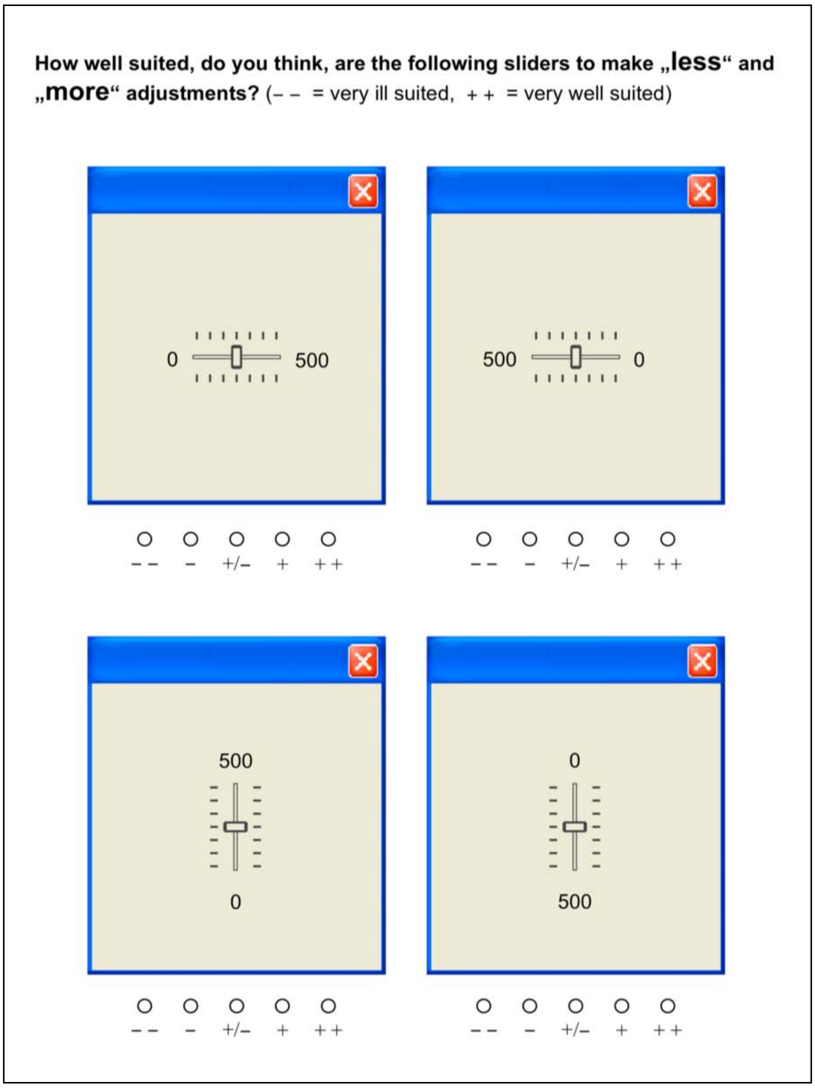


Figure 5.8. Sample page of the questionnaire used in experiment 2 for collecting suitability judgments for the target domain of quantity (translation from German)

5.3.2 Results

Slider labels that are metaphor-congruent are expected to lead to less errors, faster response times, and higher subjective suitability ratings than incongruent slider labels. These effects are expected to be smaller or not apparent in the condition with the horizontal sliders. An exception is the target domain of quantity for which stronger effects are expected based on the metaphor MORE IS RIGHT – LESS IS LEFT.

The following sections present the results in more detail. Again, the low difficulty of the task led to a ceiling effect regarding error rates. Overall accuracy was as high as 97.2%. The majority of 61.3% of the participants made no errors at all (averaged across conditions). As this ceiling effect compromises the meaningfulness of statistical results, no further statistics of the accuracy data are reported. For all other data tables see appendix A.2.

Vertical Sliders

Response times (as indicators of mental efficiency) show the expected effect of mapping: congruent mappings had significantly lower response times than incongruent mappings, $F(1,39) = 42.70$, $p < .001$. Further, there were significant effects for target domain, $F(2,78) = 176.17$, $p < .001$, and for the interaction between target domain and mapping, $F(1.60,62.21) = 8.43$, $p < .01$. The effect of target domain reflects differences between the short number labels in the target domain of quantity and the longer verbal labels in the other target domains. The interaction effect is due to larger response time differences for the target domain UP-DOWN than for the other target domains (*Figure 5.9*).

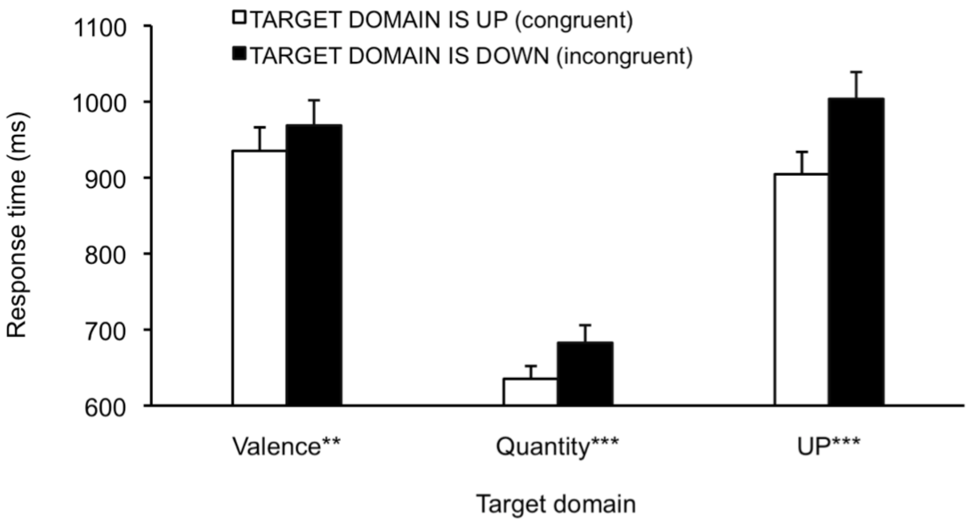


Figure 5.9. Experiment 2, vertical sliders: Response times for different target domains and mapping conditions. ** p < .01, *** p < .001 (Paired t-tests, Bonferroni-Holm corrected)

Single comparisons show metaphor-congruent mappings are significantly faster than metaphor-incongruent mappings in all target domain conditions: valence,

$t(39) = -2.47$, $p<.01$, $d = .17$; quantity, $t(39) = -4.60$, $p < .001$, $d = .37$; and physical UP-DOWN, $t(39) = -6.35$, $p < .001$, $d = .48$ (*Figure 5.9*).

Suitability judgments (as indicators of satisfaction) were higher for congruent mappings and lower for incongruent mappings, $F(1,39) = 282.76$, $p < .001$. There were also significant effects for target domain, $F(1.52,59.07) = 3.57$, $p < .05$, and for the interaction between target domain and mapping, $F(2,78) = 14.18$, $p < .001$.

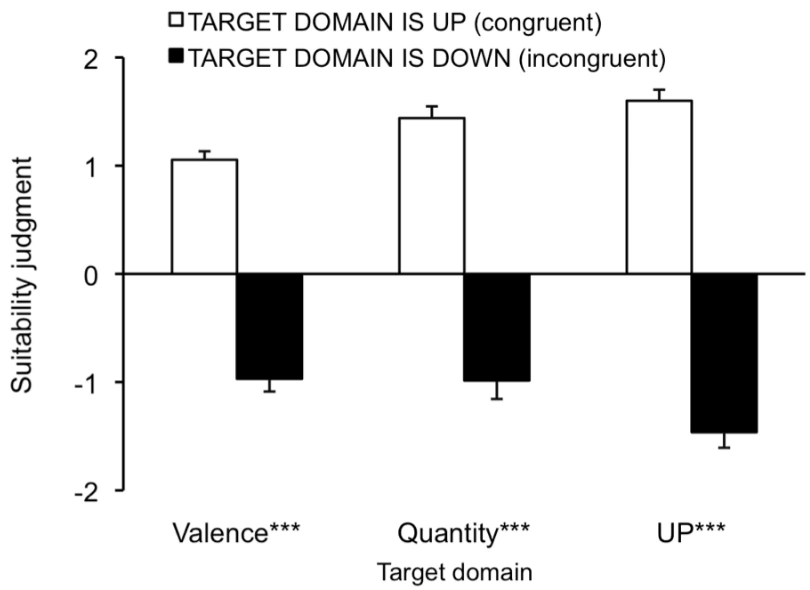


Figure 5.10. Experiment 2, vertical sliders: Suitability judgments for different target domains and mapping conditions. *** p < .001 (Paired t-tests, Bonferroni-Holm corrected)

Single comparisons revealed that metaphor-congruent mappings were rated more suitable than metaphor incongruent mappings in all target domains: valence, $t(39) = 13.91$, $p < .001$, $d = 3.20$; quantity, $t(39) = 10.99$, $p < .001$, $d = 2.68$; and physical UP-DOWN, $t(39) = 16.30$, $p < .001$, $d = 3.89$ (*Figure 5.10*).

Horizontal Sliders

Response times (mental efficiency) show no differences between metaphor-congruent and metaphor-incongruent mappings, $F(1,39) < 1$. There was a significant effect of target domain, $F(2,78) = 309.39$, $p < .001$. The interaction between target domain and mapping was not significant $F(2,78) < 1$.

Single comparisons show no significant differences between congruent and incongruent mappings in the target domains of valence, quantity, and physical UP-DOWN, $t(39) = 1.33$, $t(39) = -0.49$, and $t(39) = 0.57$, respectively.

Regarding suitability judgments (satisfaction), congruent mappings were generally preferred over incongruent mappings, $F(1,39) = 113.72$, $p < .001$. The interaction effect between target domain and mapping was significant, $F(2,78) = 42.61$, $p < .001$, indicating larger differences in the target domain of quantity than for other target domains. There was also a significant effect for target domain, $F(2,78) = 28.24$, $p < .001$.

Single comparisons revealed that the hypotheses were confirmed for all target domains. Congruent mappings were preferred over incongruent mappings for valence, $t(39) = 7.30$, $p < .001$, $d = 1.79$; quantity, $t(39) = 13.29$, $p < .001$, $d = 2.93$; and physical UP-DOWN, $t(39) = 8.38$, $p < .001$, $d = 1.86$ (*Figure 5.11*).

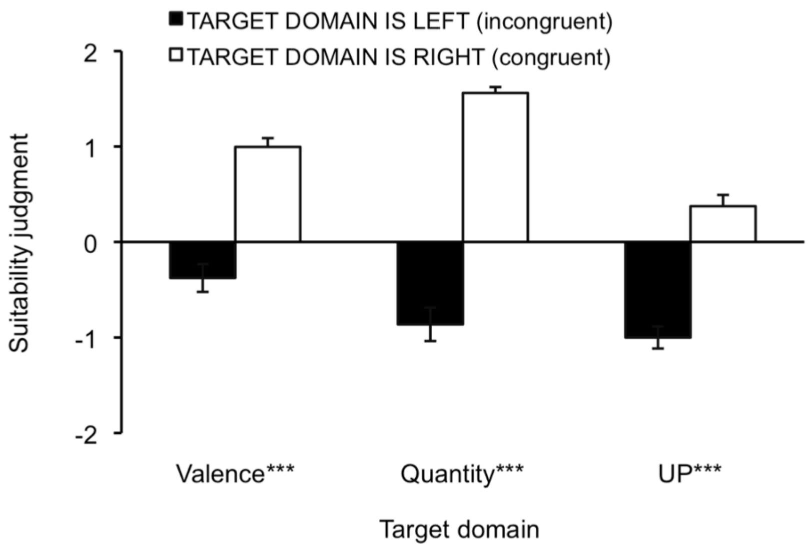


Figure 5.11. Experiment 2, horizontal sliders: Suitability judgments for different target domains and mapping conditions. *** p < .001 (Paired t-tests, Bonferroni-Holm corrected)

5.3.3 Discussion

The slider experiment replicated the findings of the button experiment on a general level. All hypotheses made for the vertical condition were confirmed. User interfaces that were designed in congruence with the metaphors POSITIVE VALENCE IS UP – NEGATIVE VALENCE IS DOWN and MORE IS UP – LESS IS DOWN proved to be more intuitive to use than user interfaces that violated these metaphors (Table 5.6). Gains in mental efficiency could be as large as 7% for the metaphors, compared to a gain of 10% for the physical UP-DOWN control condition. Satisfaction ratings showed a significant preference for metaphor-congruent slider labels over incongruent slider labels. Again, the effect sizes were low to medium for mental efficiency, and large for satisfaction measures.

Again, the evaluation of error rates was not possible, because the task was too easy. Accuracy rates showed a ceiling effect that precluded meaningful analyses of differences in effectiveness between conditions. Further studies should use more difficult tasks in order to gain results on this measure of intuitive use.

Table 5.6. *Summary of Predictions and Results for Vertical Sliders*

	Prediction	Results	
Target domain	Intuitive Use	Mental Efficiency	Satisfaction
Overall	TARGET DOMAIN IS UP > TARGET DOMAIN IS DOWN	TARGET DOMAIN IS UP > TARGET DOMAIN IS DOWN ☑	TARGET DOMAIN IS UP > TARGET DOMAIN IS DOWN ☑
Valence	POSITIVE IS UP > POSITIVE IS DOWN	POSITIVE IS UP > POSITIVE IS DOWN ☑	POSITIVE IS UP > POSITIVE IS DOWN ☑
Quantity	MORE IS UP > MORE IS DOWN	MORE IS UP > MORE IS DOWN ☑	MORE IS UP > MORE IS DOWN ☑
Physical UP-DOWN	UP IS UP > UP IS DOWN	UP IS UP > UP IS DOWN ☑	UP IS UP > UP IS DOWN ☑

Note. Results are marked with ☑ if the hypothesis was confirmed.

The results of the horizontal condition also confirmed the hypotheses on a global scale (Table 5.7). However, the effects are divided between the two indicators of intuitive use. There was no effect of mapping on mental efficiency (response times) but a large effect of mapping on satisfaction (suitability judgments). Does this point to a dissociation between online subconscious and offline conscious metaphor processing in the LEFT-RIGHT case? It may be then, that the convention of, for example, MORE IS RIGHT – LESS IS LEFT is consciously available to users and they consciously expect user interfaces to be designed that way. However, in online interaction these metaphorical extensions of the LEFT-RIGHT image schema have no influence on subconscious processing and mental performance. Note that this is in contradiction to the previous findings on the SNARC effect. Although this effect is well documented, it does strongly depend on cultural writing conventions. SNARC effects have not been found in cultures that write from right to left or from up to down (Dehaene, Bossini, & Giraux, 1993; Hung, Hung, Tzeng, & Wu, 2008). As the previous experiments did only include German participants, no valid conclusions can be drawn. What can be concluded so far is that the hypothesis seems to hold that LEFT-RIGHT provides a weaker source domain for metaphorical extensions than UP-DOWN.

In comparing the results of the two experiments it is striking to see how much the introduction of the SCALE image schema by using sliders instead of buttons and the neutralisation of the evaluative task context contribute to the effects of the valence and quantity metaphors. In the button experiment, no response time effects for the target domain of quantity were found in neither of the two source domain conditions UP-DOWN and LEFT-RIGHT. The slider experiment led to a significant difference in response times in the UP-DOWN condition (as was expected in the button experiment).

Table 5.7. *Summary of Predictions and Results for Horizontal Sliders*

	Prediction	Results	
Target domain	Intuitive Use	Mental Efficiency	Satisfaction
Overall	TARGET DOMAIN IS RIGHT ≥ TARGET DOMAIN IS LEFT	TARGET DOMAIN IS RIGHT = TARGET DOMAIN IS LEFT ☑	TARGET DOMAIN IS RIGHT > TARGET DOMAIN IS LEFT ☑
Valence	POSITIVE VALENCE IS RIGHT ≥ POSITIVE VALENCE IS LEFT	POSITIVE VALENCE IS RIGHT = POSITIVE VALENCE IS LEFT ☑	POSITIVE VALENCE IS RIGHT > POSITIVE VALENCE IS LEFT ☑
Quantity	MORE IS RIGHT > MORE IS LEFT	MORE IS RIGHT = MORE IS LEFT ☒	MORE IS RIGHT > MORE IS LEFT ☑
Physical UP-DOWN	UP IS RIGHT ≥ UP IS LEFT	UP IS RIGHT = UP IS LEFT ☑	UP IS RIGHT > UP IS LEFT ☑

Note. Results are marked with ☑ if the hypothesis was confirmed, and ☒ if not confirmed.

The target domain of quality/valence also showed different patterns of results in the horizontal conditions between experiments. While in the button study, users were faster with and preferred the mapping GOOD IS LEFT – BAD IS RIGHT, in the slider study they preferred the opposite mapping of POSITIVE IS RIGHT – NEGATIVE IS LEFT (there were no response time differences). It is as if the quality/valence target domain has been ‘quantified’ by adding an instance of the SCALE image schema. Positive valence now becomes mapped to more, negative valence to less quantity. In contrast, the first experiment used buttons that instantiate the CONTAINER image schema. This put the emphasis on qualitative distinctness and discrete values, deemphasising quantitative aspects.

Further support for a CONTAINER versus SCALE explanation of these results comes from the different results on the suitability judgements for the control condition with physical UP-DOWN items. In the button experiment, there was no specific preference whether UP should be mapped to LEFT or RIGHT. In the slider experiment, users preferred a mapping from UP to RIGHT. Again, the SCALE image schema might have acted to quantify physical UP-DOWN. This would be coherent with MORE IS UP – LESS IS DOWN that becomes MORE IS RIGHT – LESS IS LEFT.

Hence, a mapping from UP to RIGHT is a logically coherent solution. However, further research must reveal how much of these results is indeed rooted in changing the image-schema instances from CONTAINER to SCALE and how much is left to the evaluative context of the task under which participants operate.

Again, the alternative explanations of the results do not seem to be plausible, at least for the valence and the physical UP-DOWN conditions. Although average word lengths and frequencies did differ between the two words of a label (cf. appendix A.2.2), the horizontal and the vertical conditions show different patterns of results. In the case of suitability judgments, they even contradict each other. Labels that are preferred to be located in an UP position are rejected when they are presented in the LEFT position.

5.4 General Discussion and Conclusion

The two experiments described in this chapter have shown that image-schema theory can provide valid heuristics for user interface design. As hypothesised, metaphorical extensions of the UP-DOWN image schema were stronger than those for the LEFT-RIGHT image schema for the target domains of valence and quantity.

The experiments have shown that there is reason to believe that the metaphorical extensions of the UP-DOWN image schema found in language also have psychological reality and can be exploited in a user interface design context. However, in the LEFT-RIGHT conditions, there were effects on the chosen target domains that are not predicted by linguistic findings alone. For instance, the MORE IS RIGHT – LESS IS LEFT mapping is not found in linguistic expressions but is a common convention, e.g. in displaying numerical data using diagrammatical representations. The studies show that this convention influences how user interfaces are judged by the user. However, the effect is not readily apparent in performance data. Thus, it remains to be determined how generalisable user interface designs based on such conventions are – especially when applied to different contexts and cultures.

The comparison between the button and the slider study further showed that, depending on which other image schemas are instantiated in the user interface, spatial-to-abstract mappings can differ widely. This has been shown in the case of the target domain of quality. Using buttons, the mapping GOOD IS LEFT – BAD IS RIGHT fared better than GOOD IS RIGHT – BAD IS LEFT. Using sliders, this relationship was reversed. Moreover, UP-DOWN metaphors of the target domain of quantity gained in significance when buttons were replaced by sliders. Further research, however, needs to determine what role the context of the task plays in attaining these results.

Summarising the findings reported in this chapter, tentative guidelines may be formulated (Table 5.8). These guidelines are tentative in that further replications need to consolidate the findings. For instance, it might be important to find out whether the differences in representing valence on the LEFT-RIGHT dimension are really due to instantiations of a CONTAINER versus a SCALE image schema. Alternatively, these differences could depend on a similar differentiation between discrete and continuous representations. For instance, a SCALE image schema can be instantiated in continuous form by sliders or in discrete form by a list of radio buttons with ordered labels. Further studies also need to confirm these findings with more difficult tasks allowing the measurement of effectiveness data (i.e. error rates). Because the tasks in these studies were somewhat artificial to allow for better control of influencing variables, further studies may want to apply more realistic scenarios to investigate image-schematic metaphors in user interface design.

Table 5.8. *Tentative Guidelines Derived from the Empirical Results*

If you want to represent	Using the spatial dimension	And deploying instantiations of the image schema	Then use the mapping
Valence information	UP-DOWN	CONTAINER or SCALE	POSITIVE VALENCE IS UP – NEGATIVE VALENCE IS DOWN
	LEFT-RIGHT	CONTAINER	POSITIVE VALENCE IS LEFT – NEGATIVE VALENCE IS RIGHT
		SCALE	POSITIVE VALENCE IS RIGHT – NEGATIVE VALENCE IS LEFT
Quantity information	UP-DOWN	CONTAINER or SCALE	MORE IS UP – LESS IS DOWN
	LEFT-RIGHT	CONTAINER or SCALE	MORE IS RIGHT – LESS IS LEFT

What is the value of these findings to the domain of human-technology interaction? In general, they have shown that with image-schema theory one can go beyond traditional approaches to spatial-to-abstract mappings captured in the notion of population stereotypes. In particular, three aspects are important. First, image-schema theory points out previously undocumented population stereotypes. In these experiments, the metaphorical extensions of the UP-DOWN and LEFT-RIGHT image schemas extend the documented MORE IS UP / RIGHT – LESS IS DOWN / LEFT population stereotypes to a number of other target domains (i.e. virtue, quality) subsumed under the metaphor POSITIVE VALENCE IS UP – NEGATIVE VALENCE IS DOWN.

Second, by providing response time data for the existing quantity stereotypes these experiments go beyond the traditional method of devising population stereotypes from subjective preferences only. Performance data should be established as standard measures in the investigation of population stereotypes, because they can be important indicators of intuitive use and usability.

Finally, image-schema theory has been validated for a specific set of population stereotypes. So far, population stereotypes have only been enlisted without explanations of their origin and of their representation in the users' minds. Image-schema theory can provide these explanations and opens up a large host of many more potential stereotypes that await empirical testing.

The studies so far focussed on investigating the effects of different metaphorical extensions in isolation. However, in practice several target domains of the same image schema often apply that may be contradicting in their predictions. This issue is addressed in the next chapter using the example of conflicting metaphorical extensions of the image schema NEAR-FAR.

6 Conveying Abstract Information with NEAR-FAR

This chapter introduces the NEAR-FAR image schema and two of its metaphorical extensions. Under specific circumstances, these extensions make contradicting predictions for intuitive use. Two experiments investigate the effects of this contradiction.

6.1 NEAR-FAR in Human Experience

The NEAR-FAR image schema is a topological abstraction, related to the spatial proximity and distance of entities or sets of entities. Entities being NEAR or FAR are fundamental aspects of our spatial experiences in the world. Objects can be NEAR either to each other (the object view) or to the observer (the observer view) so that interactions between them or with them are possible. Normally, only limited interaction possibilities exist with objects in FAR positions.

Several metaphorical mappings are associated to the object view and the observer view of the NEAR-FAR image schema. Two of them are SIMILAR IS NEAR – DIFFERENT IS FAR and CONSIDERED IS NEAR – NOT CONSIDERED IS FAR. Both mappings are relevant to user interface design.

6.1.1 Object View

The object view of NEAR-FAR can be defined in terms of spatial regions in which common interactions between objects are possible (NEAR) or not (FAR; Baldauf, 1997). Also, the gestalt law of proximity emerges from the object view (Koffka, 1935). Objects that are NEAR to each other tend to be grouped and tend to be seen as belonging together or even as belonging to the same category.

Similar things in the natural environment tend to occur close together in space: flowers, trees, rocks, insects, gregarious animals. Also, on a geographical scale, similarity decreases with distance. World regions that are close to each other tend to be more similar, e.g. in climate, landscape, linguistic dialects, flora and fauna (Montello, Fabrikant, Ruocco, & Middleton, 2003; Nekola & White, 1999). In the built environment, buildings tend to look the same within quarters of a city and look different with increasing spatial distance between quarters. People with similar interests and values tend to meet in the same spaces. Artefacts like tools in toolboxes and crockery in cupboards are ordered in a way that reflects the similarities and differences between them.

The correlation of similarity and proximity is so ubiquitous that physical closeness stands in for similarity and distance for difference. This is also expressed in language, for example *A and B are close, but they are by no means identical. The difference between A and B is vast. There's a long way between*

Paul Newman and Woody Allen. (Lakoff, Espenson, Goldberg, & Schwartz, 1991, p.60; Lakoff & Johnson, 1999, p.51).

First studies on the psychological reality of the conceptual metaphor SIMILAR IS NEAR – DIFFERENT IS FAR were conducted outside as well as within a user interface design context. In both approaches the physical distance between objects is varied and it is determined how this variation influences similarity judgments.

Outside a user interface design context, it was investigated how similarity ratings varied with the distance of stimuli presented on a computer screen (Casasanto, 2009b). Participants made similarity judgments of pairs of abstract nouns like *grief-justice* or *memory-hope*. The results revealed that when the words of one pair were presented physically closer together the concepts were rated more similar than when they were presented further apart. However, in a second experiment, when participants were asked to rate the similarity of faces, the results were reversed. NEAR faces were judged as more different than FAR faces. Casasanto (2009b) explains these findings with the facilitation of visual comparison in the NEAR condition. When faces are presented close together in a perceptual similarity task, more differences between faces are revealed than when the faces to be compared are presented with a larger distance between them.

The reversal of effects between the perceptual judgment task and the conceptual judgment task was replicated using a set of line drawings of common objects (Casasanto, 2009b, experiment 3). When participants judged the functional similarity of the depicted objects (a conceptual judgment), results were consistent with the metaphor SIMILAR IS NEAR. When the task was to judge the visual similarity, the results were reversed and contradicted the predictions of the SIMILAR IS NEAR metaphor.

The SIMILAR IS NEAR metaphor was also successfully exploited as a method for determining the similarity of objects (Goldstone, 1994). Participants arranged items on a computer screen in a way that the proximity of the objects was proportional to their similarity. This task was done much faster than more traditional similarity ratings on a questionnaire and the data obtained by this method highly correlated with traditional ratings.

Within a user interface design context the SIMILAR IS NEAR metaphor was investigated in the domain of information visualisation (Montello et al., 2003). Participants were told that each point on a point display represented a document (*Figure 6.1*). Three of the document points were labelled with *A*, *1*, and *2*. Participants rated the similarity of *A* and *1* and compared it to their rating of the similarity of *A* and *2*. Results indicate that closer pairs are rated as more similar than more distant pairs. Physical distance ratios (distance A-1 / distance A-2) and similarity ratings correlated r=.70.

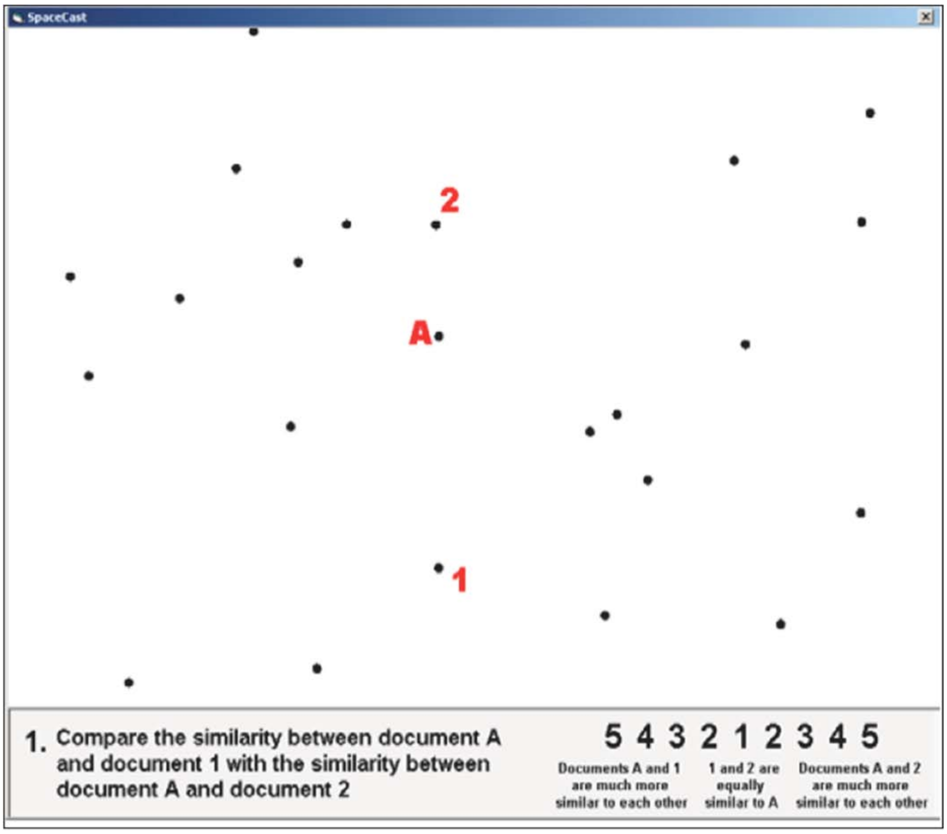


Figure 6.1. Sample screenshot of the experiment of Montello et al. (2003) that investigated the SIMILAR IS NEAR metaphor in a user interface context. The screen shows a point display, the similarity question, and the rating scale.

A limitation of the point-display study is that the distance between points was the only clue available to users to derive similarity judgments from. All displayed points looked the same and no further information about the documents was available. The evidence would be more convincing, if the source domain of the metaphor were not obligatory in judging the similarity of the stimuli. This is the case in the Casasanto (2009b) study outside the user interface context, in which all necessary information for the similarity ratings was present in the stimuli, but task-irrelevant distance variations had still an influence on these ratings.

To fully judge the contribution of stimulus distance to the degree of intuitive use, more than the effect on the judgments itself needs to be investigated. In the following experiments, all three outcomes of intuitive use are investigated, i.e. effectiveness, mental efficiency, and satisfaction.

6.1.2 Observer View

The observer view of the NEAR-FAR image schema is grounded in the sensorimotor experience of objects being within (NEAR) or without reach (FAR) of the observer. This differentiation between reaching zones is supported by experimental findings. When people estimate distance, objects located at the same distance seem closer when the objects can be reached than when they are out of reach (Witt, Proffitt, & Epstein, 2005). Neuropsychological studies found that in patients with hemispatial neglect, NEAR and FAR space can be independently impaired (Cowey, Small, & Ellis, 1994; Halligan & Marshall, 1991).

The observer view entails that NEAR objects can be inspected in detail so that more information can be gathered about them than about FAR objects. In addition, NEAR-FAR may not only express the distance to our body but also the distance from the current centre of attention. Accordingly, HERE-THERE was proposed as a synonym of the NEAR-FAR image schema, illustrated by the sentence "see the star HERE (where I am pointing with the index finger), and the other one THERE (farther to the left)" (Krzeszowski, 1997, p. 118).

The observer perspective on NEAR-FAR is the basis of the metaphorical extension CONSIDERED IS NEAR – NOT CONSIDERED IS FAR. As being NEAR often means 'being within reach' of the observer, NEAR objects can be seized, grasped, inspected, and manipulated. This correlation of physical distance with mental access of physical objects is metaphorically extended to abstract domains, for instance when talking about ideas or action possibilities, for example in *My companion put it to me that an initiative must now be taken. For the Kaszubes and Poles of Danzig, Poland was a distant idea only. Far be it from us to condone tax evasion. He did come close to going on pension.* (BNC, 2007).

As mentioned before, the image schema NEAR-FAR can also be used to describe the distance from the current focus of attention (Krzeszowski, 1997). Items that are NEAR the current focus of attention will have a greater chance to be considered than items that are FAR from the current focus of attention.

The metaphorical mapping CONSIDERED IS NEAR – NOT CONSIDERED IS FAR is already a central guideline for user interface design. Designing from the task- and user-centred viewpoint means organising the user interface in a way that information needed to solve a task is presented in physical proximity (Mayhew, 1992). This idea also underlies the proximity compatibility principle. The principle specifies that when a task requires the mental integration of multiple sources of information, performance will be best when that information is displayed in close proximity (Wickens & Carswell, 1995). In other words, the degree of *processing proximity* should be matched by the degree of *display proximity*. Processing proximity defines the extent to which two information sources are used within the same task, for example, whether they have to be integrated or compared. That is, the use of these information sources is in temporary closeness. Display proximity is based originally on spatial NEAR-FAR relations between display components. However, and confusingly, it can also mean that user interface elements are similar in terms of colour or information coding (Wickens & Carswell, 1995).

In order to draw a more precise line between the metaphors used, in the following studies the proximity compatibility principle is considered in its original form, i.e.

as an expression of the metaphor CONSIDERED IS NEAR (in contrast to the implied secondary reading of CONSIDERED IS SIMILAR). Unfortunately, this distinction has not always been made in studies on the proximity compatibility principle, so that the evidence is difficult to judge.

6.2 Study 3: Single Comparisons of Displays

This study is designed in a way that the two metaphorical extensions of the NEAR-FAR image schema make the same predictions in one condition and contrasting predictions in another condition. The participant's task is to judge the similarity of values in two displays that are in different positions on a 3x3 array of displays – NEAR to or FAR from each other. The hypothesis is that these task-irrelevant NEAR-FAR relations influence the effectiveness, mental efficiency, and user satisfaction when making similarity judgments.

In line with the proximity compatibility principle, the metaphor CONSIDERED IS NEAR – NOT CONSIDERED IS FAR predicts that it is always more intuitive when the displays are NEAR to each other because they are considered together in the same task. The metaphor SIMILAR IS NEAR – DIFFERENT IS FAR predicts that it is more intuitive to place similar values NEAR to each other and different values FAR from each other than placing different values NEAR to each other and similar values FAR from each other. Therefore, if the two displays show similar values, both metaphors, CONSIDERED IS NEAR and SIMILAR IS NEAR make the same prediction. Users' similarity judgments should be faster, more accurate, and more satisfying, if the two displays are NEAR to each other than when they are FAR from each other. In the other case, when displays show different values, the two metaphors make contrasting predictions. The prediction of the CONSIDERED IS NEAR metaphor remains the same: users make faster and more accurate judgements and are more satisfied when the two displays are NEAR to each other than when they are FAR from each other. The predictions of the SIMILAR IS NEAR metaphor change. Users now should be faster, more accurate, and more satisfied, when the displays are FAR from each other than when they are NEAR to each other (Table 6.1).

Table 6.1. *Summary of the Hypotheses in Experiment 3*

	Predictions of the metaphor	
Similarity condition	CONSIDERED IS NEAR	SIMILAR IS NEAR
Similar values	NEAR > FAR	
Different values	NEAR > FAR	NEAR < FAR

Note. The greater than (>) and smaller than (<) signs refer to the expected degree of intuitive use when the two displays are in NEAR vs. FAR positions relative to one another.

The reasoning behind these predictions is that the varying distance between the displays activates the NEAR-FAR image schema that is associated with both metaphors. Hence, both metaphors become activated. In the condition in which both metaphors make the same prediction, a metaphor-congruent user interface will immediately trigger a response. A metaphor-incongruent user interface requires inhibiting these pre-activated metaphors and requires building new mappings appropriate for solving the task. Therefore, the actual response will take longer than for the metaphor-congruent user interface.

In the condition in which the metaphors make contrasting predictions, there will be a process of contention scheduling leading to generally higher response times than in the condition in which both metaphors make the same prediction. In this case there are two possible outcomes, depending on the relative strength of activation of the metaphors. (1) If both metaphors are activated with the same strength, no further differences between user interfaces are expected, because the user interface is always congruent with one of the metaphors and incongruent with the other. The resulting facilitation and inhibition effects are expected to cancel each other out. (2) If one metaphor is more strongly activated than the other, the task should be solved faster with user interfaces that are congruent with the stronger metaphor than user interfaces that are incongruent with the stronger metaphor. A metaphor can get stronger activation by lowering its activation threshold. The threshold can be lowered permanently due to stronger associations in experience. On a temporary basis, the threshold could be lowered due to the circumstances of the task, for example a specific goal. If this hypothesis is appropriate, it could be expected that the similarity judgment task temporarily lowers the activation threshold for the metaphor SIMILAR IS NEAR. Consequently, it is more activated than CONSIDERED IS NEAR and user interfaces should be more intuitive to use when they are congruent with the SIMILAR IS NEAR metaphor.

Two types of displays were used in the experiment. Displays either showed numbers or analogue pointers at different angles. The two conditions were chosen to compare conceptual judgments (comparing numbers) with perceptual judgments (comparing angles of pointers). Although conceptual metaphor theory predicts the same results for both display types, previous studies indicated that there might be differences between similarity judgments using conceptual (abstract words) and perceptual (human faces) comparisons (Casasanto, 2009b, see above). Here, the hypotheses follow conceptual metaphor theory and no differences between number and pointer displays are expected.

Participants were instructed to judge the similarity of the displayed values by means of their numerical or angular difference (in number and pointer displays, respectively). The basis of the similarity judgements was deliberately held simple to retain a degree of control over the concept of similarity the participants employed. This simple task implies a one-dimensional model of similarity processing (Shepard, 1974) instead of a more complex feature-based model of similarity (Tversky, 1977) or an account of structural alignment (Gentner & Markman, 1997). To investigate the influence of NEAR-FAR variations on similarity judgments based on more complex similarity models is left to future research.

6.2.1 Method

Participants

Twenty-four participants (14 male, 10 female) were recruited from the campus of Technische Universität Berlin. They were between 18 and 34 years old ($M = 26.6$, $SD = 4.2$). Participation took place in exchange for payment (eight Euros).

Procedure

First, participants completed a demographic questionnaire. They then received a written instruction in which they learned about the experimental procedure (see appendix A.3). They were instructed to compare pairs of display values in number and pointer displays and to indicate whether the values were similar or not. The participants were asked to judge the similarity of values in number displays by their numerical difference. The criterion for judging the similarity of pointer displays was their angular difference. Participants were not provided with a pre-defined threshold indicating the similarity or dissimilarity of displays. Participants were made aware of the range of the values (10 to 100) the displays could assume. They were instructed to respond as fast and accurate as possible.

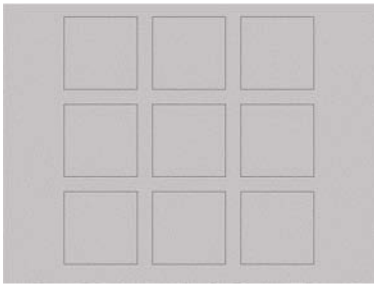
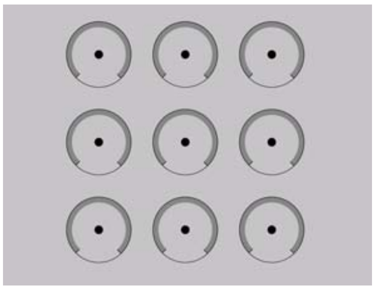
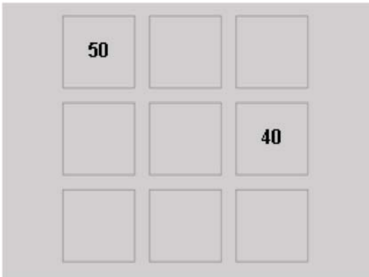
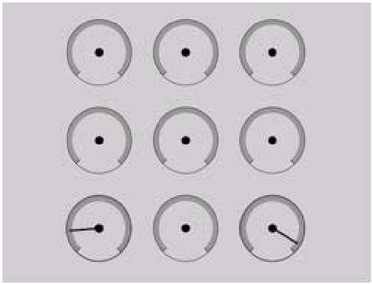
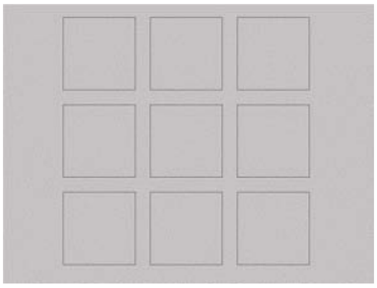
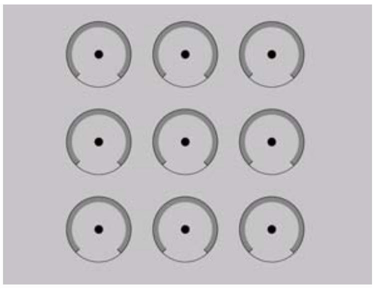
Each display condition consisted of 198 trials and was preceded by 33 practice trials (the data of which did not enter statistical analyses). Half of the participants received the number-display condition first, the other half the pointer-display condition. For each trial the procedure is described in Table 6.2.

After finishing the trials of each block (number or pointer displays), participants filled in a questionnaire containing questions about the suitability of different forms of display arrangements for judging the similarity of values. The experiment ended with an interview containing questions about the thresholds the participants used for deciding whether two displays were similar or different, about their response strategies, and about their intuitions about the purpose of the experiment. After the interview, a debriefing about the intended purpose of the experiment was given.

Material

Stimuli. The values and the positions of each pair of displays were varied independently from each other. Display values ranged between 10 and 100 in steps of 10 and were randomly assigned to trials. In the number-display condition these numerical values were directly used for presentation. In the pointer-display condition these numbers were mapped to pointer angles in 27° increments on a scale covering a total of 270° (*Figure 6.2*).

Table 6.2. *Trial Structure of Experiment 3*

Number-display condition	Pointer-display condition	Events
		1. Arrangement of empty displays shown for 500 ms.
		2. Two displays are randomly filled with values. Participants decide whether they deem the two values as similar or different. They respond with either the left or the right index finger on the left (the ‘D’) or the right button (the ‘L’) on a laptop keyboard.
		3. After the button press (or a maximum time of 5000 ms) the screen shows the arrangement of empty displays again. After 250 ms the next trial starts.

Note. Step 2: For half of the participants the ‘similar’ response was assigned to the left key and the ‘different’ response to the right key, for the other half the key assignments were reversed. Between trials participants kept their fingers over the left and right keys.

Figure 6.2. Display values used in the pointer-display condition. The far left display corresponds to a value of 10, the second to a value of 20, and so on; the last corresponds to a value of 100.

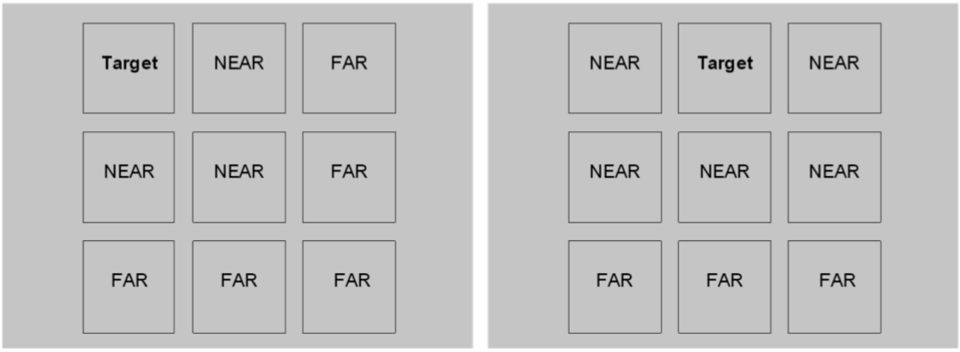


Figure 6.3. Coding of NEAR-FAR relations for possible positions of a target value in a corner (left) or at the other marginal positions (right) in the 3x3 array of displays.

The display proximity of any pair of displays was coded as NEAR when they appeared in adjacent positions. The other pairs were coded as FAR. *Figure 6.3* illustrates this coding for different positions of a target display in the 3x3 array. It shows that for a target display placed in any of the four corners, there are three possible NEAR positions and five possible FAR positions of the second display (*Figure 6.3* left). For the four marginal positions between the corners of a display there are five possible NEAR and three possible FAR conditions of the second display (*Figure 6.3* right). The number of possible NEAR and FAR relations are thus counterbalanced over all eight marginal target positions of the 3x3 array. Target positions in the centre of the array are establishing only NEAR relations. They were not included, because they would off-balance the design in favour of NEAR relations. Excluding the centre target position lead to a total source of (9-1) x 8 = 64 position pairs that were randomly assigned to displays.

Apparatus. Software and hardware for stimulus presentation and data collection were the same as in experiment 1. The 3x3 array of displays extended 201x201mm for number displays and 195x195 mm for pointer displays. The vertical and horizontal distance between the centres of adjacent displays was 72 mm in both, pointer and number displays. The average distance between the centres of NEAR displays (coded according to the possible positions in *Figure 6.3*) was 83 mm. The average distance between the centres of FAR displays was 159 mm. The viewing distance was 60 cm. For the number labels, a 20 pt Arial font was used.

Questionnaires. The same demographics questionnaire as in experiment 1 was used. Suitability judgments (as an indicator of satisfaction) were collected by a questionnaire that presented four 3x3 arrays of displays in each of which two displays showed any values. All four arrays showed the same display values but in randomly varying positions. Two of the 3x3 arrays always contained a display pair at NEAR positions, the other two a display pair at FAR positions (*Figure 6.4*). The sequence of these positions varied between questionnaire pages and was counterbalanced across subjects.

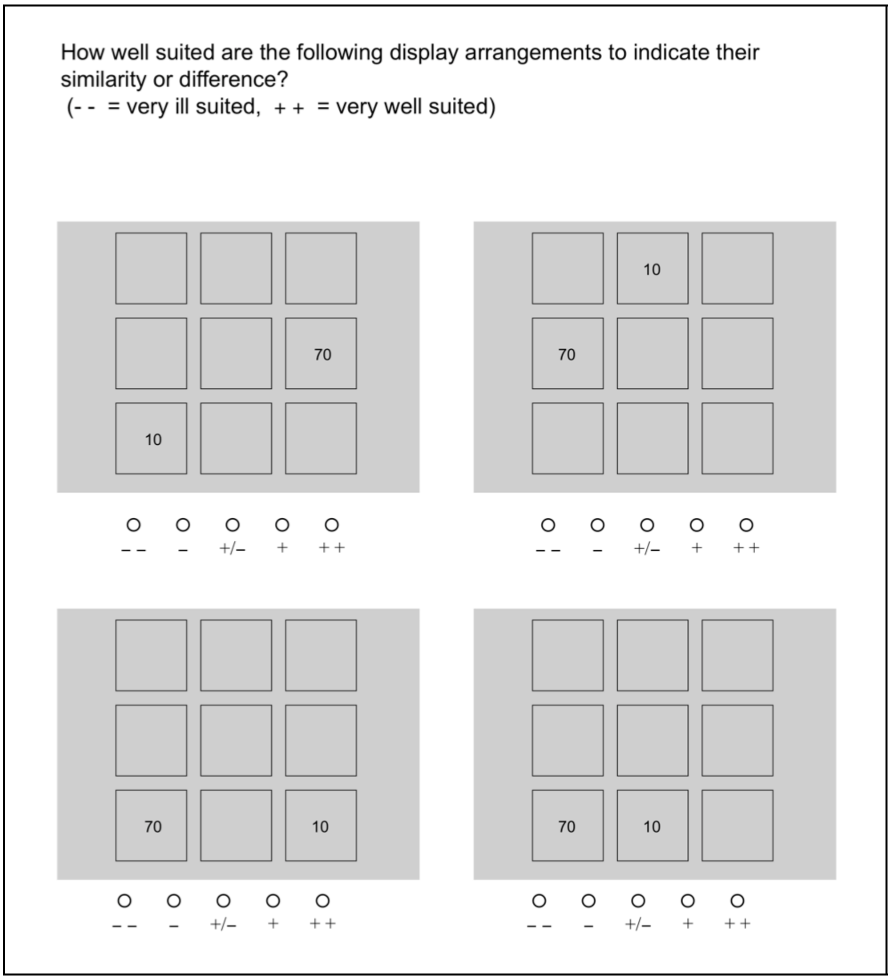


Figure 6.4. Sample page of the questionnaire for gathering suitability judgments (number-display condition, similarity type "different"; translation from German)

Three types of similarity relations of display values were used: similar (with absolute differences between display values < 30), different (with absolute differences between display values > 50), and intermediate (with absolute differences between display values between 30 and 50). Each participant received six pages of each type, a total of 18 pages. The sequence of these pages was random. They were preceded by an instruction page. There were two types of questionnaires, one for each condition of displays (pointer or number displays).

Experimental Design

A within-subjects design was used with three independent variables: display type (number displays, pointer displays), similarity ("similar" or "different" – at individual thresholds), and display proximity (NEAR or FAR). As dependant variable, the degree of intuitive use was measured. As in the previous experiments, error rates are used as an indicator of effectiveness, response times are used as indicators of mental efficiency, and suitability judgments are used as indicators of satisfaction.

For the classification of display values into similar and different values, individual similarity thresholds were determined from the logfile data. First, absolute differences between display values were computed. Then the percentage of 'similar' judgements at each level of absolute difference was determined. For example, at the levels of absolute differences of 0, 10, 20, 30, and 40, the relative frequencies of 'similar' judgements could be 100%, 97%, 100%, 33%, 0%, respectively. The threshold then was set between two adjacent levels of absolute difference in which the smaller absolute differences had a frequency of 'similar' responses above 50% and the larger absolute differences of below 50%. In the example the threshold is set between the absolute differences of 20 and 30 and numerically expressed as 25 (the mean value of 20 and 30).

Errors were computed according to these individual similarity thresholds. Response times were taken from the log file of the computerised experiment. Suitability judgments were collected with the questionnaire described above. A total of 3.6% of the data across all dependant variables were outliers that were replaced by mean values at the subject level. The procedures for data analysis were the same as in study 1.

6.2.2 Results

The experiment was set up to compare two metaphorical extensions of the NEAR-FAR image schema. The first, CONSIDERED IS NEAR – NOT CONSIDERED IS FAR, maintains that if two displays are compared for their similarity, then they should always be placed NEAR each other. Placing them FAR from another will lead to higher response times, higher error rates, and lower suitability judgments. The second metaphor, SIMILAR IS NEAR – DIFFERENT IS FAR, predicts that the optimal placement of displays is NEAR for displays of similar values and FAR for displays of different values. Thus, similar displays in FAR positions and different displays in NEAR positions should result in lower performance and lower suitability ratings.

In terms of overall results, the metaphor CONSIDERED IS NEAR predicts a main effect of display proximity. The metaphor SIMILAR IS NEAR predicts an interaction effect between display proximity and similarity.

Results contain mixed evidence. There was no clear support in favour of one of the metaphors. The metaphor CONSIDERED IS NEAR is partly supported by response time and subjective data. SIMILAR IS NEAR is partly supported by error data. The following sections contain the detailed results. More details can be found in appendix A.3.

Similarity Thresholds

Subjects were asked to rate similarity in terms of differences between display values. They were free to set their individual difference thresholds. Below this threshold they rated pairs of displays as similar to each other and above this threshold they rated pairs of displays as different from each other. The mean similarity threshold over all participants in the number-display condition was 27 (*SD* = 11.0, *min* = 15, *max* = 55). In the number-display condition, individual thresholds did not differ between the NEAR and FAR conditions, $t(23) < 1$, n.s. In the pointer-display condition, the mean threshold was 73° (*SD* = 28.4°, *min* = 41°, *max* = 122°). Similarity thresholds were higher when the pointer displays were presented NEAR each other than when presented FAR from another, $t(23) = -1.74$, $p < .05$, as would be predicted by the SIMILAR IS NEAR metaphor. In other words, participants judged pointer displays more similar to each other, when they were NEAR than when they were FAR.

Empirically determined thresholds corresponded well with the threshold values the participants reported in the interview after the experiment. The correlation for number displays was $r(24) = .96$, $p < .001$, and for pointer displays $r(23) = .74$, $p < .001$. The lower value for pointer displays reflects that judging and reporting angular differences is more difficult than judging and reporting arithmetic differences. As this confirms the validity of the empirically derived individual thresholds, these were used in the following analyses to differentiate 'different values' from 'similar values'.

Error Rates (Effectiveness)

Generally, error rates were not different between the NEAR and FAR conditions – except for different values in number displays. Here, when the displays were NEAR, more errors were made as when they were FAR. This result follows a pattern predicted by the SIMILAR IS NEAR metaphor (*Figure 6.5*).

The different and similar value conditions were determined by individual difference thresholds (see section 6.2.1). Error rates were computed from the amount of 'similar' responses in the different-values condition and from the amount of 'different' responses in the similar-values condition. Error rates were significantly lower for different values than for similar values, $F(1,23) = 10.85$, $p < .01$. Display proximity had no significant main effect, $F(1,23) < 1$. Neither was the interaction between display proximity and similarity significant, $F(1,23) < 1$.

Single comparisons show significant differences between the NEAR and FAR condition in the number-display condition for different values, $t(23) = 2.83$, $p < .05$, $d = .70$, but not for the other conditions, all $t(23) < 1$. The drop in error rate in the significant condition is 61%; the effect size *d* is medium to large.

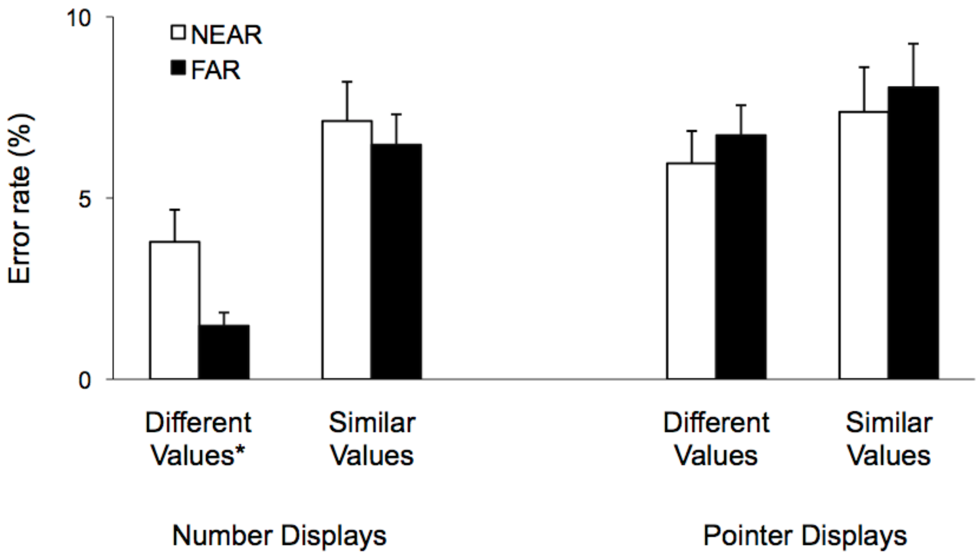


Figure 6.5. Error rates for NEAR-FAR relations of pairs of displays (number or pointer displays) that either showed different or similar values (experiment 3). * $p < .05$ (Paired t-tests, Bonferroni-Holm corrected)

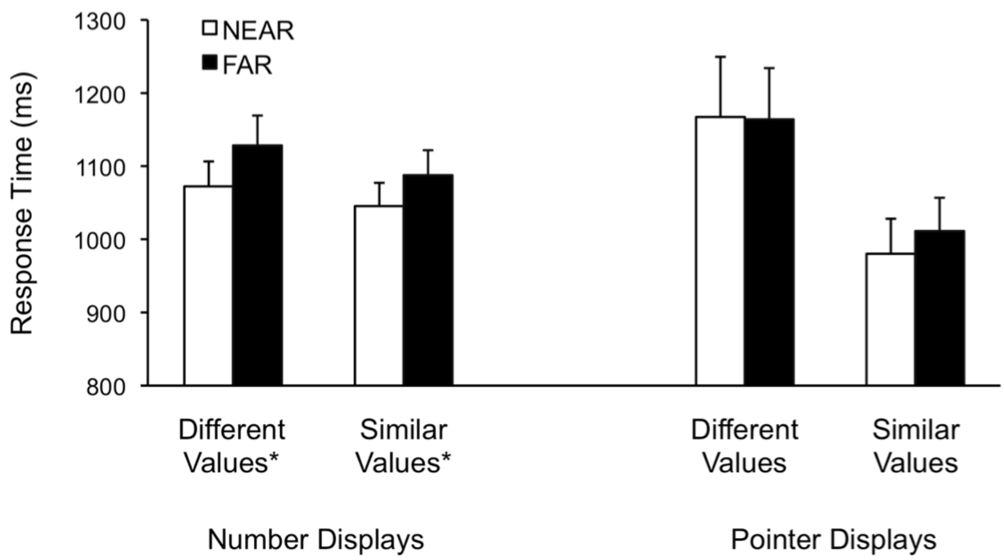


*Figure 6.6***.** Response times for NEAR-FAR relations of pairs of displays (number or pointer displays) that either showed different or similar values (experiment 3). * $p < .05$, (Paired t-tests, Bonferroni-Holm corrected)

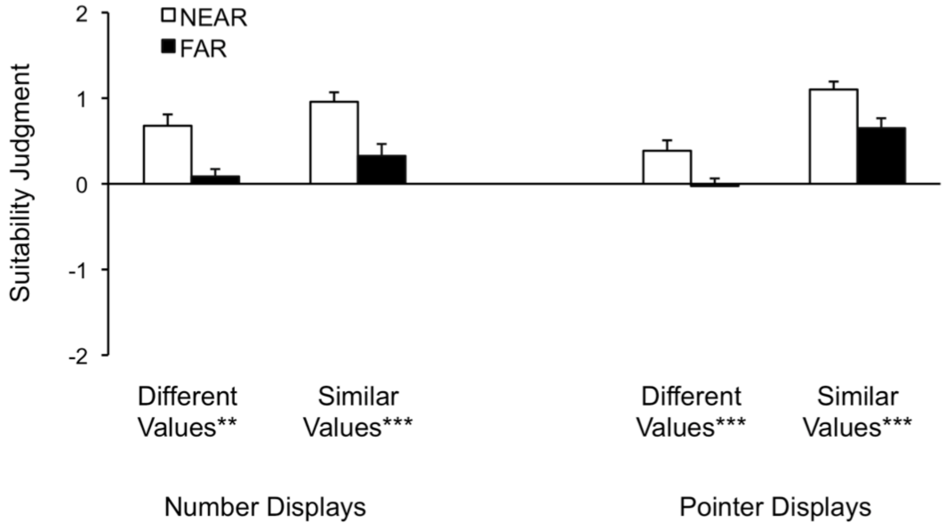


Figure 6.7. Suitability ratings for NEAR-FAR relations of pairs of displays (number or pointer displays) that either showed different or similar values (experiment 3). ** p < .01, *** p < .001 (Paired t-tests, Bonferroni-Holm corrected)

Response Times (Mental Efficiency)

The response times follow the pattern predicted by CONSIDERED IS NEAR, but only in the number-display condition. No decision for or against any hypothesis could be made within the pointer-display condition. Response times were the same with NEAR and FAR distances of pointer displays (*Figure 6.6*).

There was a main effect for display proximity in that participants made faster decisions with NEAR than with FAR displays, $F(1,23) = 7.34$, $p < .05$. The interaction between display proximity and similarity was not significant, $F(1,23) < 1$. Response times were significantly higher for different values than for similar values, $F(1,23) = 19.42$, $p < .001$.

Single comparisons show significant differences between the NEAR and FAR conditions for number displays: for different values, $t(23) = 2.52$, $p < .05$, $d = .30$, and for similar values: $t(23) = 2.44$, $p < .05$, $d = .26$. No effect of display distance on response times was found in the pointer-display conditions, neither for different values $t(23) < 1$, nor for similar values, $t(23) = 1.29$. The gains in processing speed in the number-display conditions are 5% and 4%, respectively; the effect sizes are small to medium.

Suitability Judgments (Satisfaction)

Generally, the suitability judgments were higher for the NEAR condition regardless of whether the display values were similar or not. This result follows the pattern predicted by the CONSIDERED IS NEAR metaphor (*Figure 6.7*).

Regarding display proximity, the NEAR condition was judged more suitable than the FAR condition, $F(1,23) = 33.41$, $p < .001$. The interaction between display proximity and similarity was not significant, $F(1,23) < 1$.

Single comparisons showed higher ratings for NEAR display positions than for FAR positions in the number-display condition for different values, $t(23) = 3.33$, $p < .01$, $d = 1.08$, and for similar values, $t(23) = 4.52$, $p < .001$, $d = 1.03$. Higher ratings for NEAR display positions were also obtained in the pointer-display conditions, for different values, $t(23) = 3.80$, $p < .001$, $d = .79$, and for similar values, $t(23) = 5.46$, $p < .001$, $d = .88$.

6.2.3 Discussion

Is one metaphor stronger than the other? The results of this study are ambiguous. The findings on the individual similarity thresholds support the metaphor SIMILAR IS NEAR. Displays were judged more similar when they were NEAR to each other than when they were FAR from another. However, this effect was only apparent in the pointer condition.

Both metaphors, SIMILAR IS NEAR and CONSIDERED IS NEAR make the same predictions on intuitive use in the condition with similar values. These predictions were confirmed by the satisfaction data in both display conditions, but not by the effectiveness data. The mental efficiency data matched the prediction in the number, but not in the pointer-display condition.

Response times were generally higher in the condition with different values in which both metaphors make contrasting predictions. The detailed results do not lead to a clear conclusion in favour of one of the metaphors (Table 6.3). Regarding performance data, differences can only be found in the number-display condition. Here, the effectiveness results support the SIMILAR IS NEAR metaphor, but the mental efficiency results support the CONSIDERED IS NEAR metaphor. Errors were higher but response times were shorter when displays with different values were placed NEAR each other than when they were placed FAR from another.

The results on satisfaction generally supported the metaphor CONSIDERED IS NEAR that is in consistence with the proximity compatibility principle described above. Thus, while participants in an offline task find a solution better that maps task-relevant items NEAR each other, their online performance data does not fully agree.

Where does this seemingly inconsistent dominance of either the one or the other metaphor come from in the performance data? Looking more closely at the time and error results, it becomes obvious that participants traded speed for accuracy. The faster they were, the more errors they produced. In both, pointer and number-display conditions, response times are lower and error rates are higher in general for the similar condition than for the different condition (*Figure 6.5* and *Figure 6.6*). The pattern is repeated within the different condition in the number displays:

NEAR displays show higher speed and higher error rates than FAR displays. But note the asymmetry: the speed difference is just 5%, the difference in error rates is 61% (also reflected in effect sizes, $d = .30$ and $.70$, respectively).

Table 6.3. *Summary of the Predictions and Results in Experiment 3*

<table>
<tr><th></th><th></th><th colspan="2">Predictions</th><th colspan="2">Results</th></tr>
<tr><th>Similarity condition</th><th>Intuitive Use component</th><th>SIMILAR IS NEAR</th><th>CONSIDERED IS NEAR</th><th>Number display</th><th>Pointer display</th></tr>
<tr><td rowspan="3">Similar values</td><td>Effectiveness</td><td colspan="2">NEAR > FAR ☒</td><td>NEAR = FAR</td><td>NEAR = FAR</td></tr>
<tr><td>Mental Efficiency</td><td colspan="2">NEAR > FAR (☑)</td><td>NEAR > FAR</td><td>NEAR = FAR</td></tr>
<tr><td>Satisfaction</td><td colspan="2">NEAR > FAR ☑</td><td>NEAR > FAR</td><td>NEAR > FAR</td></tr>
<tr><td rowspan="3">Different values</td><td>Effectiveness</td><td>NEAR < FAR (☑)</td><td>NEAR > FAR ☒</td><td>NEAR < FAR</td><td>NEAR = FAR</td></tr>
<tr><td>Mental Efficiency</td><td>NEAR < FAR ☒</td><td>NEAR > FAR (☑)</td><td>NEAR > FAR</td><td>NEAR = FAR</td></tr>
<tr><td>Satisfaction</td><td>NEAR < FAR ☒</td><td>NEAR > FAR ☑</td><td>NEAR > FAR</td><td>NEAR > FAR</td></tr>
</table>

Note. Confirmation of a hypothesis is marked with ☑, rejection with ☒. Results that only partly support a hypothesis are indicated by (☑).

A practitioner may be tempted to decide for the metaphor SIMILAR IS NEAR. The large drop in errors and the higher effect size that goes with it could motivate this decision. However, CONSIDERED IS NEAR might be more convincing to those practitioners that value subjective data over performance data. Clearly, deriving guidelines on these results is on shaky grounds. More conclusive evidence has to be gathered.

One reason for the speed-accuracy trade-off could have been that the task was not difficult enough for the participants. They had to make binary decisions on a predefined and rather clear-cut definition of similarity. The manipulation of display distance did not yield large effects because participants had enough processing capacity left to cope with any possible interference of distance with similarity judgments. Thus, the task was not able to elicit clear effects favouring either the one or the other metaphor (especially in the pointer-display condition). Wickens and Hollands (2000) discuss that the speed-accuracy trade-off may vanish with more difficult tasks, so that slower responses go along with higher error rates. According to them, tasks are more difficult in this respect when they pose higher memory loads and/or exhibit poorer stimulus evidence. To avoid a speed-accuracy trade-off, the next experiment (study 4) makes the task more difficult by increasing the memory load. It is expected that study 4 will produce more decisive results on the question of which metaphor is able to explain the findings better.

6.3 Study 4: Double Comparisons of Displays

In this study the difficulty of the comparison task was enhanced using double comparisons. The same types of stimuli were used as in study 3 (number and pointer displays). Similar to the Montello et al. (2003) study described above, participants now had to decide which of two pairs of displays showed more similar values than the other pair. Each display pair consisted of the same target display (*Figure 6.8*, display A) and one of two comparison displays (displays 1 and 2).

In the 'similar' task condition, participants indicated which of the two display values (1 or 2) was more similar than the other compared to the target display value (A). Participants responded by indicating the location of the more similar display on the 3x3 array of displays. In the 'different' task condition participants indicated which of the two display values (1 or 2) was more different than the other compared to the target display value (A). Participants then responded by indicating the location of the different display. The distance of the comparison displays 1 and 2 to the target display A varied, but one display was always NEAR the target display while the other display was FAR from the target display in the critical trials.

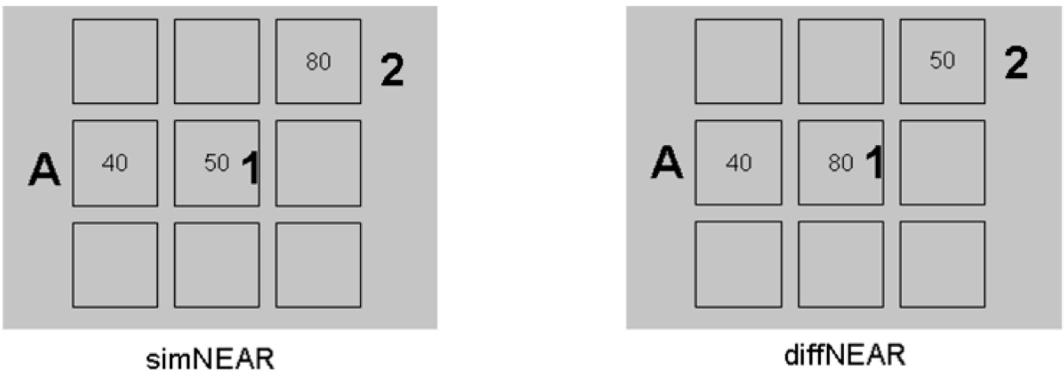


Figure 6.8. Basic components of the task in experiment 4: target display (A) and comparison displays (1 and 2) in two NEAR-FAR conditions. In the condition simNEAR the similar display was NEAR to the target display and the different display was FAR from the target display. In the condition diffNEAR the different display was NEAR to the target display and the similar display was FAR from the target display

Table 6.4. *Summary of the Hypotheses in Experiment 4*

	Hypotheses of the metaphor	
Task condition	CONSIDERED IS NEAR	SIMILAR IS NEAR
Find similar displays	simNEAR > diffNEAR	
Find different displays	simNEAR < diffNEAR	simNEAR > diffNEAR

Note. simNEAR denotes an experimental condition in which the similar display is NEAR to and the different display is FAR from the target display. diffNEAR denotes the experimental condition in which the different display is NEAR to and the similar display is FAR from the target display. The greater than (>) and smaller than (<) signs refer to the expected degree of intuitive use (i.e. effectiveness, mental efficiency, and satisfaction).

In the previous study participants had to decide whether two values are similar or different and then responded by pressing a 'similar' or 'different' button. The current set-up was more difficult as there were more cognitive steps necessary to solve the task and memory load was higher. In the previous experiment participants had to (1) determine the difference between displays A and B, then (2) decide on an answer (similar or different), and (3) remember the correct key assignment before responding. In the current experiment participants had to determine two differences: between displays A and 1 and between displays A and 2 (steps 1 and 2). After remembering the assigned task (to look for the more similar or the more different display) (step 3), they compared the differences (step 4) and decided which difference is smaller or larger, respectively (step 5). In order to give a response they also had to remember the location of the display corresponding to the right answer (step 6).

Note that the distances between displays A and 1 as well as A and 2 are not important in solving the similarity task. Yet, image-schema theory predicts that metaphorical extensions of the NEAR-FAR image schema will be automatically activated by distance cues and will interfere with task solving, thus affecting response times, errors, and suitability judgements. Again, the metaphors SIMILAR IS NEAR and CONSIDERED IS NEAR make different predictions as to whether the similar displays being NEAR or FAR are more intuitive to use (Table 6.4).

Assuming that there is always a pair of displays that is more similar than the other pair and both pairs are at different distances, the following predictions hold. The metaphor SIMILAR IS NEAR – DIFFERENT IS FAR predicts that an arrangement in which the more similar display is NEAR display A and the more different display is FAR from display A is always more beneficial than when the more different display is NEAR and the more similar display is FAR.

The metaphor CONSIDERED IS NEAR – NOT CONSIDERED IS FAR predicts that the display that answers the task is more beneficially placed NEAR the target display than FAR from it. More specifically, when the task is to find the more similar display, an arrangement in which the similar display is NEAR the target display and the more different display is FAR from the target display should be more beneficial than when the more different display is NEAR and the more similar display is FAR. When the task is to find the more different display, it should be more beneficial to place the more different display NEAR the target display and the more similar display in a FAR position than to place the more similar display NEAR and the more different display FAR (Table 6.4). Both metaphors predict the same pattern of results in the task condition ‘similar’ and contrasting results in the task condition ‘different’.

6.3.1 Method

Participants

Twenty-four participants (12 male, 12 female) were recruited from the campus of Technische Universität Berlin. They were between 19 and 52 years old ($M = 27.3$, $SD = 6.8$). Participation took place in exchange for payment (eight Euros).

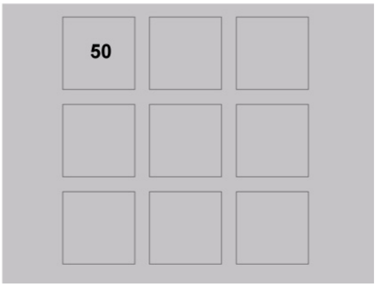
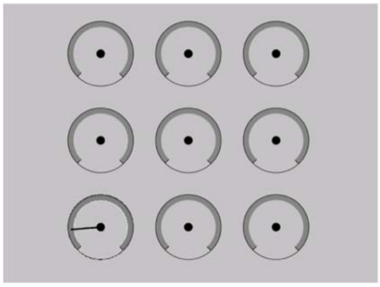
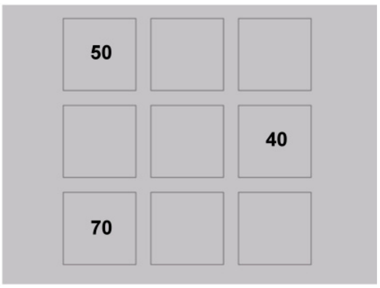
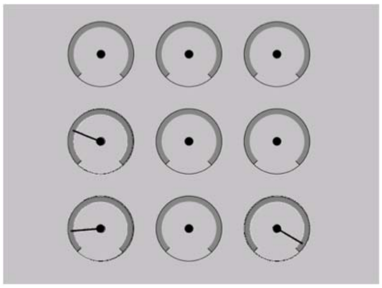
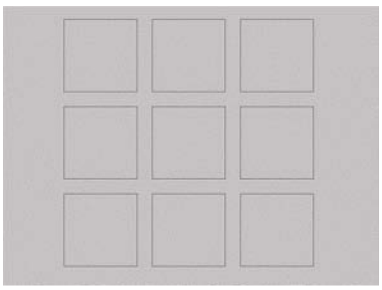
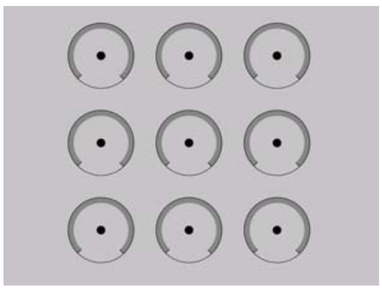
Procedure

First, participants completed a demographic questionnaire. They then received a written instruction in which they learned about the experimental procedure (see appendix A.4). They were instructed to compare pairs of display values of number and pointer displays. For each type of display there were two task blocks: one asking “Which displays are more similar?” the other “Which displays are more different?” Each task block consisted of 60 trials and was preceded by 20 practice trials. Again, the participants were instructed to base their similarity judgments of number displays on numerical differences and of pointer displays on angular differences. Participants were also instructed to respond as fast and accurately as possible.

The procedure for each trial is shown in Table 6.5. The response time was measured from the beginning of the first button press (step 2) to the release of this button (step 4). This arrangement made it possible to measure only the time needed for the cognitive steps of solving the task, without a confoundation by the time needed to steer the finger from the location of the first button to the location of the second button press (the actual response).

After finishing the trials of each condition (number or pointer displays), participants filled in a questionnaire containing questions about the suitability of different forms of display arrangements for judging display similarity and dissimilarity. The experiment ended with a short interview about their response strategies and about their intuitions about the purpose of the experiment. After the interview participants received a debriefing about the intended purpose of the experiment.

Table 6.5. *Trial Structure of Experiment 4*

Number-display condition	Pointer-display condition	Events
 OR		1. The task is presented for 2000 ms.
		2. An arrangement of displays is shown. A random position is filled with a random value (the target display). Participants press the button on a 3x3 labelled keypad that corresponds to the display's location.
		3. As long as this button is pressed (or for a maximum time of 5000 ms), the comparison displays are shown. Participants decide which of the two displays best answers the question.
		4. As soon as the button is released, the screen shows nine empty display fields. Participants respond by pressing a button that corresponds to the location of the correct display. Then the next trial starts.

Note. Step 2: The correct button to press in this example is the upper left button in the number-display condition or the lower left button in the pointer-display condition. Step 4: When answering the question "Which displays are more similar?", the correct button is the middle-right button in the number-display condition or the middle-left button in the pointer-display condition. When answering the question "Which displays are more different?", the lower-left button or the lower-right button has to be pressed, respectively.

Material

Stimuli. The values and the positions of each triplet of displays were varied independently from each other. Display values ranged between 10 and 100 in steps of 10. Each triplet of display values consisted of a *target* value, a value that was *similar* to the target, and a value that was *different* from the target. The absolute difference between the target value and the more different value was always greater than the absolute difference between the target and the more similar value. More similar values could be either the same as the target value or differ from it (maximum possible absolute difference = 80). Obeying these rules resulted in a number of 430 value triplets, a pool from which value triplets were randomly drawn and assigned to trials. In the number-display condition these values were directly used for presentation, in the pointer-display condition these numbers were mapped to 10 images of pointer angles in 27 degree increments on a scale covering 270 degrees (as in study 3, *Figure 6.2*).

These triplets of display values were shown at randomly varying positions. For each of the nine positions of the target value, 8x7 positions of the other two values were possible, resulting in 504 possible position triplets. Any display pairs with adjacent positions were coded as NEAR; the other pairs were coded as FAR (as in study 3, see *Figure 6.3*). Again, target positions in the centre of the array were not included, because they would off-balance the design in favour of NEAR relations. This led to a total of (504 – 8 x 7 =) 448 position triplets.

Table 6.6. *Displays Positions, Their Roles in Experiment 4, Their Expected, and Actual Frequencies*

Role of item in experiment	Display(s) NEAR to the target display	Display(s) FAR from the target display	Expected frequency, relative and (absolute)	Actual frequency in the experiment
Experimental (simNEAR)	Similar	Different	26.8% (120)	26.1%
Experimental (diffNEAR)	Different	Similar	26.8% (120)	27.3%
Filler (bothNEAR)	Similar, Different	–	23.2% (104)	22.7%
Filler (bothFAR)	–	Similar, Different	23.2% (104)	23.9%
Total			100% (448)	100%

Value triplets were randomly assigned to position triplets. As the more similar display could be NEAR or FAR to the target display and also the more different display could be NEAR to or FAR from the target display, four types of position relations were possible (Table 6.6). Of these, only two types, in which the display position of the similar target differed from the display position of the different target (NEAR-FAR), were used to validate the hypotheses (see above, called simNEAR and diffNEAR). The others, in which both the more similar and the more different display were NEAR or FAR the target display, served as filler items. This was done to occlude the experiment's hypotheses from the participants. Table 6.6 also shows the frequency of these position relations achieved by random selection from the source of 448 position triplets in the actual experiment. Expected and actual frequencies did not differ, $\chi^2(3) = 3.47$, $p = .325$.

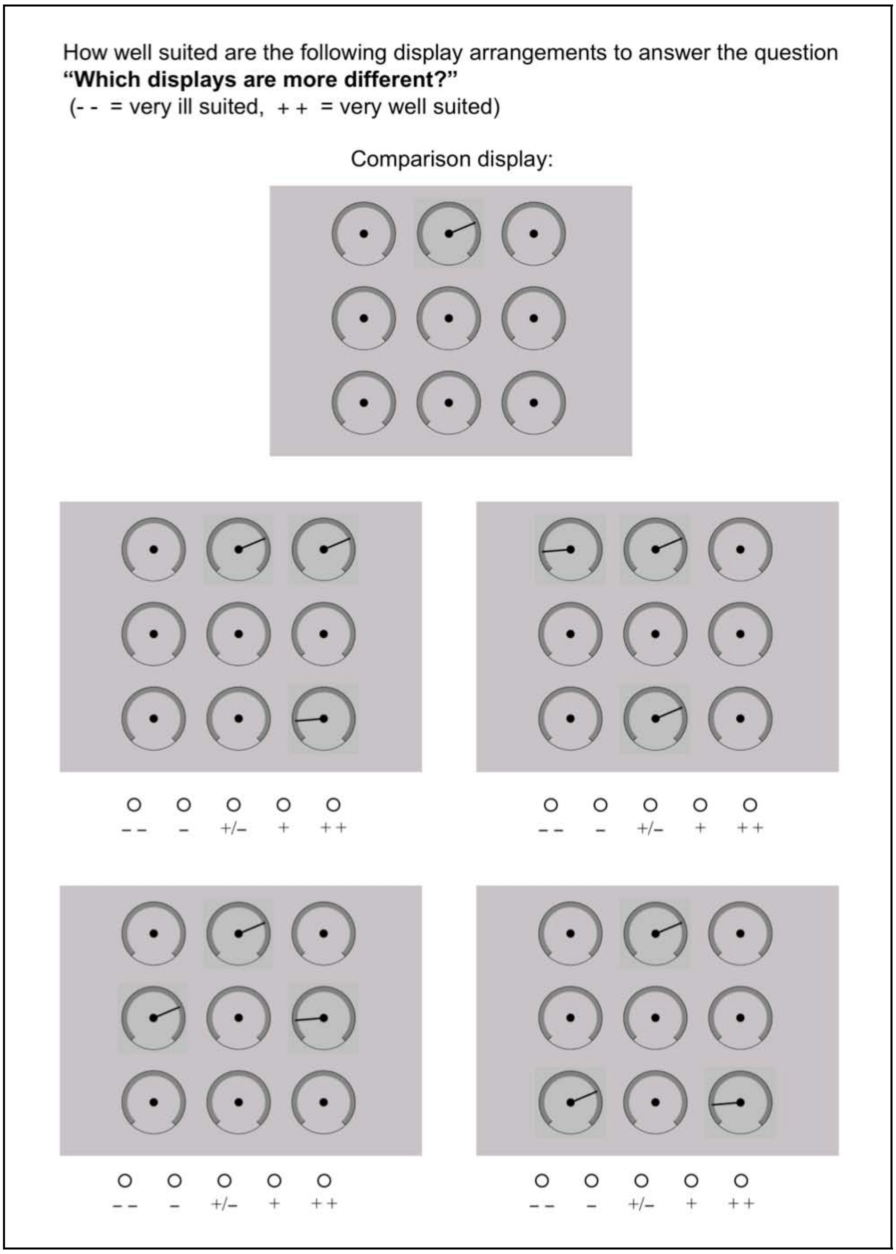


Figure 6.9. Sample page from the questionnaire in experiment 4 (pointer-display condition, task type 'different'; translation from German)

Apparatus. Experimental set-up, software, and hardware for stimulus presentation and data collection were the same as in experiment 1. The size of and the distances within the 3x3 array of displays were the same as in experiment 3.

Questionnaires. The same demographics questionnaire as in experiment 1 was used. Suitability judgments (as indicators of satisfaction) were collected with a questionnaire that presented on each page first a 3x3 array of displays with a target value for reference and then four 3x3 arrays each containing the target value at a fixed position and the same similar and different display values in different arrangements: simNEAR, diffNEAR, bothNEAR, and bothFAR (Figure 6.9). Again, the bothNEAR and bothFAR arangements were used as filler items. The sequence of these positions varied and was counterbalanced across subjects.

After finishing the trials of each of the two display conditions (pointer and number displays) participants received a 19-page questionnaire. It consisted of an instruction (one page) and two blocks of tasks: eight questionnaire pages related to the similarity task ("Which displays are more similar?") and eight pages related to the difference task ("Which displays are more different?"). The two blocks were each preceded by a page containing the task for the next block. The sequence of these tasks in the questionnaire was counterbalanced across subjects.

Experimental design

A within-subjects design was used with three independent variables: display type (number displays, pointer displays), task (similar: "Which displays are more similar?" and different: "Which displays are more different?"), and display arrangement (simNEAR or diffNEAR). The sequence of display types was counterbalanced between subjects and the sequence of tasks within a display type condition was randomly assigned.

As dependant variable, degree of intuitive use was measured. As in the previous experiments, error rates are used as an indicator of effectiveness, response times are used as indicators of mental efficiency, and suitability judgments are used as indicators of satisfaction. Error rates and response times were taken from the log file of the computerised experiment. Suitability judgments were collected with the questionnaire described above. A total of 3.8% of the data across all dependent variables were outliers that were replaced by mean values at the subject level. The procedures for data analysis were the same as in study 1.

6.3.2 Results

The experiment was set up to compare the data against two competing predictions of metaphorical extensions of the NEAR-FAR image schema. The first, SIMILAR IS NEAR, maintains that displays showing more similar values should be placed NEAR the target display and that displays showing more different values should be placed FAR from the target display. Arranging displays like this should lead to faster response times, fewer errors, and higher suitability judgments. The other metaphor, CONSIDERED IS NEAR, predicts that the optimal placement of displays that are considered for solving a task is NEAR. Thus, when asking for the more different display, the more different display should be placed NEAR the target display. When asking for the more similar display, the more similar display should be placed NEAR the target display in order to enhance performance and subjective ratings.

In terms of overall results, the metaphor SIMILAR IS NEAR predicts a main effect of display arrangement. The metaphor CONSIDERED IS NEAR predicts an interaction effect between display position and task.

Results regarding the performance data support the metaphor SIMILAR IS NEAR. The analysis of the subjective data did not contribute to a decision between both metaphors. The following sections (and appendix A.4) present details of the results.

Manipulation Check

The experiment was designed to provide the participants with a more difficult task in order to eliminate the speed-accuracy trade-off observed in the previous experiment. To check whether this goal was reached, overall performance rates were compared between both studies. The results indicate that performance with the new task was as high as with the old task in the number-display condition, but deteriorated significantly in the pointer-display condition.

In the number-display condition, response times did not significantly differ between experiment 3 and experiment 4, $t(46) < 1$. Accuracy rates even slightly increased (from 94% to 97%), $t(46) = 2.17$, $p < .05$, $d = .85$. In the pointer-display condition, response times increased significantly from an average of 1102 ms in experiment 3 to 1967 ms in experiment 4, $t(46) = 9.53$, $p < .001$, $d = 2.75$. Accuracy rates dropped from 92% in experiment 3 to 88% in experiment 4, $t(28.55) = -1.73$, $p < .05$, $d = .48$. It follows that the manipulation of the task difficulty was more successful for pointer displays than for number displays.

Error Rates (Effectiveness)

Results show that error rates largely follow the pattern predicted by the metaphor SIMILAR IS NEAR (*Figure 6.10*). A main effect for display arrangement was found in that simNEAR arrangements of displays generally led to less errors than diffNEAR displays, $F(1,23) = 22.86$, $p < .001$. The interaction between display position and task was not significant, $F(1,23) < 1$. Further, the interaction between display type and display arrangement was significant, $F(1,23) = 10.83$, $p < .01$. Accuracy differences between the conditions simNEAR and diffNEAR were larger for pointer displays than for number displays.

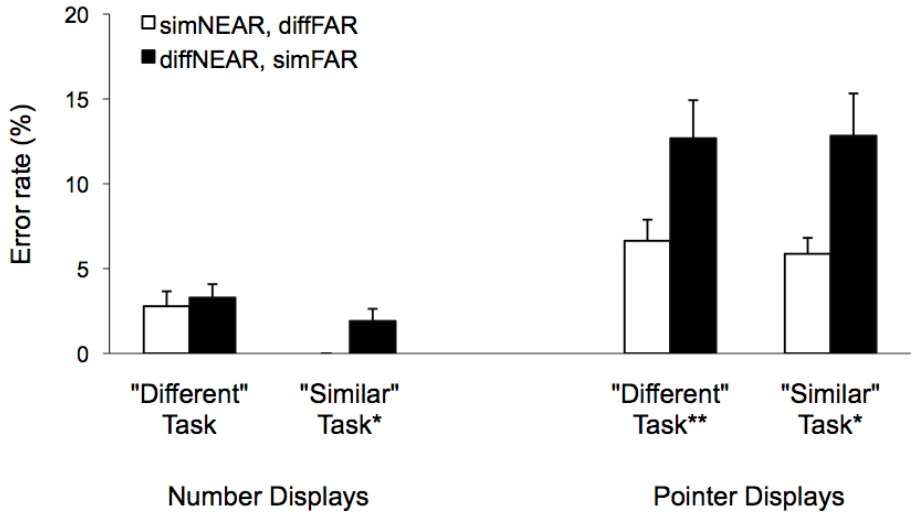


Figure 6.10. Error rates in experiment 4. The task was to either indicate the more different display or the more similar display. * p< .05, ** p < .01 (Paired t-tests, Bonferroni-Holm corrected). No errors were made in the number display, 'similar' task, simNEAR condition.

Single comparisons in the number-display condition show lower error rates in the simNEAR condition than in the diffNEAR condition only for the similar task, $t(23) = 2.68$, $p < .05$, $d = .77$, but not for the different task, $t(23) < 1$. The decrease in error rates for the similar task is 100% because no errors were made in the simNEAR condition. Single comparisons in the pointer-display condition show lower error rates for the simNEAR than for the diffNEAR condition for the similar task, $t(23) = 2.60$, $p < .05$, $d = .76$, as well as for the different task $t(23) = 3.57$, $p < .01$, $d = .68$. Error rates dropped by 54% and 48%, respectively.

Response Times (Mental Efficiency)

Results show that the response times follow the pattern predicted by the metaphor SIMILAR IS NEAR (*Figure 6.11*). There was a main effect for display arrangement in that simNEAR arrangements of displays generally were judged faster than diffNEAR displays, $F(1,23) = 28.24$, $p < .001$. The interaction between display arrangement and task was not significant, $F(1,23) = 1.06$. Only the interaction between display type and display arrangement was significant, $F(1,23) = 6.24$, $p < .05$. Response time differences between the conditions simNEAR and diffNEAR were larger for pointer displays than for number displays. Furthermore, response times in the different-task condition were generally higher than response times in the similar-task condition, $F(1,23) = 67.12$, $p < .001$.

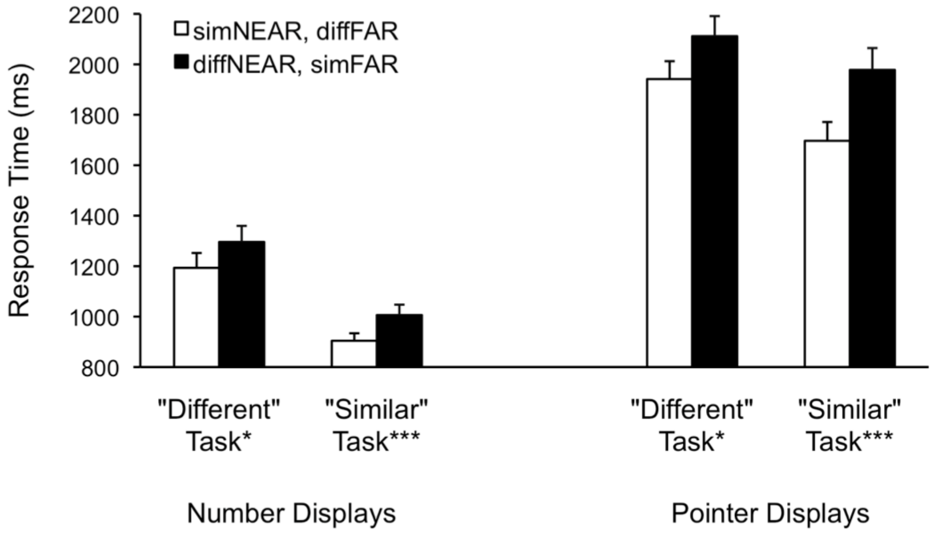


Figure 6.11. Response times in experiment 4. The task was to either indicate the more different display or the more similar display. * p < .05, *** p < .001 (Paired t-tests, Bonferroni-Holm corrected).

Single comparisons in the number-display conditions reveal significantly faster responses in simNEAR display arrangements than in diffNEAR display arrangements for the different task, $t(23) = 1.96$, $p < .05$, $d = .34$, and for the similar task, $t(23) = 4.84$, $p < .001$, $d = .58$. Single comparisons in the pointer-display conditions reveal significant advantages of the simNEAR display arrangements over the diffNEAR display arrangements for the different task, $t(23) = 2.30$, $p < .05$, $d = .46$, and for the similar task, $t(23) = 4.10$, $p < .001$, $d = .71$. Gains in processing speed are 8%, 10%, 8%, and 14%, respectively.

Suitability Judgments (Satisfaction)

Subjective results do support the predictions of both metaphors when the task was to find the more similar display. Both metaphors predict an advantage of the simNEAR condition over the diffNEAR condition (*Figure 6.12*). The results on the different task are inconclusive. At most there is a very slight advantage of the metaphor CONSIDERED IS NEAR in the number-display condition (found in the ANOVA, but not in the single comparisons).

A main effect for display position was found in that simNEAR display arrangements were generally judged more suitable than diffNEAR display arrangements, $F(1,23) = 14.25$, $p < .001$. Suitability judgements differed between the two display types, with number displays generally being judged more suitable than pointer displays, $F(1,23) = 4.33$, $p < .05$. The interaction between task and display arrangement was also significant, $F(1,23) = 8.55$, $p < .01$.

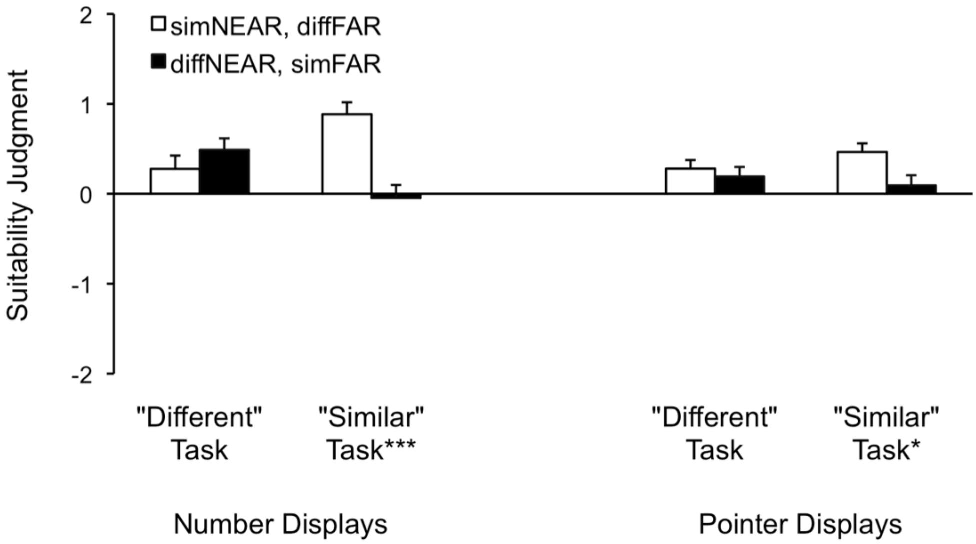


Figure 6.12. Subjective suitability ratings in experiment 4. The task was to either indicate the more different display or the more similar display. * p < .05, *** p < .001 (Paired t-tests, Bonferroni-Holm corrected).

Single comparisons show significant advantages of the simNEAR over the diffNEAR display arrangements only for the similar task in the number-display condition, $t(23) = 4.54$, $p < .001$, $d = 1.38$, and in the pointer-display condition, $t(23) = 2.58$, $p < .05$, $d = .74$. For the different task the differences between display arrangements were not significant, neither in the number condition, $t(23) = 1.01$, nor in the pointer condition, $t(23) < 1$.

6.3.3 Discussion

The experiment was designed to play two metaphorical extensions of the NEAR-FAR image schema off against each other. The metaphor SIMILAR IS NEAR proposes that when the more similar display is NEAR the target display, it is more advantageous than when it is FAR from the target display. The same prediction is made by the metaphor CONSIDERED IS NEAR when the task is to judge which display is more similar to the target display. When the task is to judge which display is more different from the target display, however, then CONSIDERED IS NEAR holds the reverse prediction. Positioning the more different display NEAR the target display is expected to be more advantageous than positioning the more similar display NEAR the target display. The results were more in favour of the metaphor SIMILAR IS NEAR than of the metaphor CONSIDERED IS NEAR, at least in the performance data (Table 6.7). Furthermore, in the different-task condition in which both metaphors contradict each other, response times were higher than in the similar-task condition – as would be predicted by contention scheduling.

Table 6.7. *Summary of the Predictions and Results in Experiment 4*

		Predictions		Results	
Task condition	Intuitive Use component	CONSIDERED IS NEAR	SIMILAR IS NEAR	Number display	Pointer display
Find similar displays	Effectiveness	simNEAR > diffNEAR ☑		simNEAR > diffNEAR	simNEAR > diffNEAR
	Mental Efficiency	simNEAR > diffNEAR ☑		simNEAR > diffNEAR	simNEAR > diffNEAR
	Satisfaction	simNEAR > diffNEAR ☑		simNEAR > diffNEAR	simNEAR > diffNEAR
Find different displays	Effectiveness	simNEAR < diffNEAR ☒	simNEAR > diffNEAR (☑)	simNEAR = diffNEAR	simNEAR > diffNEAR
	Mental Efficiency	simNEAR < diffNEAR ☒	simNEAR > diffNEAR ☑	simNEAR > diffNEAR	simNEAR > diffNEAR
	Satisfaction	simNEAR < diffNEAR ☒	simNEAR > diffNEAR ☒	simNEAR = diffNEAR	simNEAR = diffNEAR

Note. Confirmation of a hypothesis is marked with ☑, rejection with ☒. Results that only partly support a hypothesis are indicated by (☑).

As CONSIDERED IS NEAR is congruent with the proximity compatibility principle by Wickens and Carswell (1995), the results imply that the principle is not valid in the context of this experiment. The results are better explained by a mapping that has rarely been discussed in the ergonomics literature. The question arises whether the SIMILAR IS NEAR metaphor is permanently stronger than the CONSIDERED IS NEAR metaphor. In reviewing the evidence on the proximity compatibility principle, Wickens and Carswell (1995) found that the literal interpretation of display proximity in terms of spatial distance often has only weak effects, compared to a metaphorical interpretation of 'proximity' in terms of similar appearance. While this could speak for a permanent weakness of CONSIDERED IS NEAR, a temporary effect cannot be ruled out. As the task was about judging the similarity of displays, the target domain of similarity was already activated in the users' mind. Hence, the metaphor SIMILAR IS NEAR was activated more strongly and subsequently had a bigger effect on the judgment task. The task dependency of metaphor salience was also shown in an experiment by Schubert (2005). In a situation in which both metaphors contradicted each other, the competition between GOOD IS UP and POWER IS UP was decided by the task given to the participants. When participants rated the valence of stimuli on a screen, the metaphor GOOD IS UP was more salient. When they rated the power of the stimuli, the metaphor POWER IS UP was more salient (cf. chapter 3.3.1). Taking this into account, the results of experiment 4 may be better explained in terms of task dependency of the effects of conceptual metaphor. Further research needs to clarify whether changing the task also shifts the pattern of activation of different metaphors.

Another point to note is that the performance effects were not equally strong between the two display conditions, though the same tendencies applied. This could have been a result of the relatively small sample size (N = 24), lacking the power to yield significant effects for effectiveness in the number-display condition. Also, it could be an effect of differing task difficulty – pointer displays were more difficult to judge than number displays. This can be concluded from the data in the manipulation check above and from what the participants said in the interview. Effects are also expected to be more pronounced with greater distances between displays – other studies have used wall-projected images instead of laptop computer screens (Montello, et al., 2003).

6.3.4 Additional Analysis of the Filler Items

A second explanation of why the metaphor CONSIDERED IS NEAR (i.e. the proximity compatibility principle) failed could lie in the design of the experiment itself. The experimental trials were designed in a way that one of the two comparison displays was always NEAR the target display and the other was always FAR. The predictions of the metaphor CONSIDERED IS NEAR were formulated in a way that it should be beneficial when the comparison display that solves the task is NEAR the target display. Thus, the direct comparison of the predictions of both metaphors within a single condition became possible.

The proximity compatibility principle, however, is stricter, because it states that *all* displays that are needed in the task should be placed NEAR each other. More precisely, it predicts that placing *both* comparison displays NEAR the target display should be more beneficial in terms of performance and suitability ratings than placing them *both* FAR from the target display. This prediction should hold without regard to task or display type.

To check whether this more conservative interpretation of CONSIDERED IS NEAR is valid in the given context of the experiment, the filler trials of the experiment were analysed (cf. Table 6.6). Again, the independent variables were display type (number and pointer display), task (to indicate the more similar or the more different display) and display position (both displays NEAR to or both displays FAR from the target display) as within-subject factors.

Concerning the error rates, there was no significant overall difference between the condition in which both displays were NEAR the target display and the condition in which both displays were FAR from the target display, $F(1,23) = 2.56$. The interaction between display position and task was not significant, $F(1,23) < 1$. None of the single comparisons between display positions (bothNEAR and bothFAR) reached significance (see appendix A.4.6).

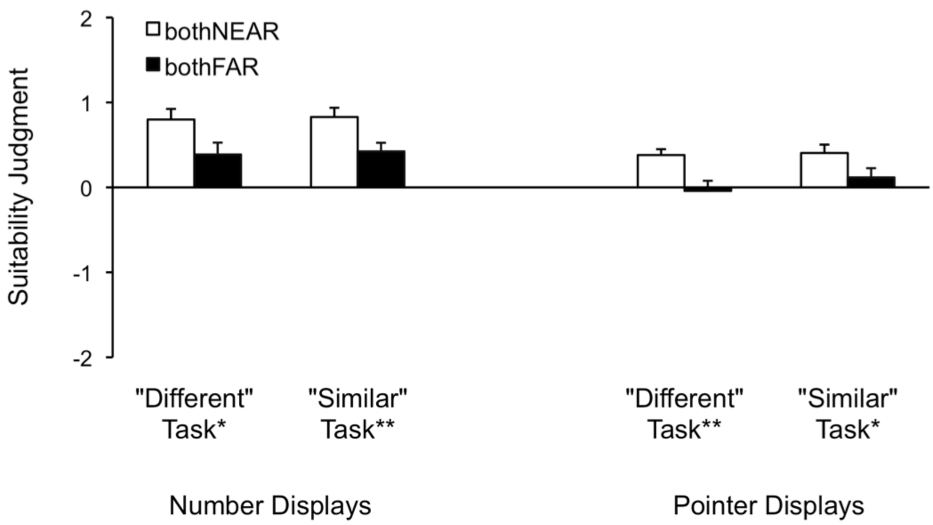


Figure 6.13. Subjective suitability ratings for the filler items in experiment 4. Conditions are compared in which both comparison displays were either NEAR the target display (bothNEAR) or both were FAR from the target display (bothFAR). * p < .05, ** p < .01 (Paired t-tests, Bonferroni-Holm corrected)

With regard to response times, there was no significant overall difference between the condition in which both displays were NEAR the target display and the condition in which both displays were FAR from the target display, $F(1,23) = 4.23$. The interaction between display position and task was not significant, $F(1,23) < 1$. None of the single comparisons reached significance (see appendix A.4.6).

As to the subjective suitability ratings, subjective data dissociated from performance data. Display arrangements in which both comparison displays were NEAR the target display were judged more suitable than display arrangements in which both comparison displays were FAR from the target display, $F(1,23) = 11.98$, $p < .01$ (*Figure 6.13*). The interaction between display position and task was not significant, $F(1,23) < 1$.

Single comparisons show significant differences between the NEAR and FAR conditions for number displays: for the different task, $t(23) = 2.49$, $p < .05$, $d = .63$, and for the similar task, $t(23) = 3.37$, $p < .01$, $d = .80$. Significant differences between the NEAR and FAR conditions were also found for pointer displays: for the different task, $t(23) = 3.02$, $p < .01$, $d = .87$, and for the similar task, $t(23) = 1.89$, $p < .05$, $d = .55$.

In summary, the performance results of the filler items did not confirm the hypotheses posed by the metaphor CONSIDERED IS NEAR. However, the predictions of the metaphor were confirmed by the subjective suitability judgments. Thus, even a more conservative interpretation of the CONSIDERED IS NEAR metaphor did not confirm the metaphor's predictions in the performance data. Following a less conservative or a more conservative interpretation of CONSIDERED IS NEAR enhanced neither the effectiveness nor the mental efficiency of interaction. Only the satisfaction was enhanced with the strict interpretation (but not with the less strict one). This indicates a strong conscious opinion, but also shows a lesser relevance to the influence of the metaphor in online product interaction, at least in the given task setting.

6.4 General Discussion and Conclusion

In experiments 3 and 4 several predictions of image-schema theory were investigated using two metaphorical extensions of the NEAR-FAR image schema. First, in the condition in which both metaphors make the same prediction, it was found that the metaphor-congruent user interface is more intuitive to use than the metaphor-incongruent user interface. The effects were most pronounced in the satisfaction data and were more stable in the performance data with higher mental workload (experiment 4) than with lower mental workload (experiment 3).

Second, in the condition in which the metaphors make contrasting predictions, participants showed slower responses than in the condition in which the metaphors make the same predictions. This could be due to a process of contention scheduling taking place when two metaphors are activated at the same time and when they have different behavioural consequences. Although the data in both experiments seem to confirm this hypothesis, alternative explanations are possible. Due to the specific design of these experiments, the contrasting-metaphor conditions overlapped with judging differences. If difference judgments are harder to perform than similarity judgments, this effect could be sufficient to explain the response time differences. Disentangling these effects needs to be addressed in further studies.

Third, it was predicted that when both metaphors have the same activation strength, no differences between NEAR and FAR conditions should be observed, because facilitation and inhibition effects will cancel each other out. If one metaphor is activated more strongly than the other, the results should favour this stronger metaphor. While the results of study 3 are difficult to interpret in this respect because of a speed-accuracy trade-off, the performance data of study 4 points to a stronger SIMILAR IS NEAR metaphor. In contradiction to the proximity compatibility principle (CONSIDERED IS NEAR), placing the more different displays further away from the target display yielded up to 14% faster and 48% less error-prone performance. An additional analysis of the filler items also did not confirm a stronger contrasting CONSIDERED IS NEAR mapping in the performance data. The subjective data, however, showed that CONSIDERED IS NEAR is readily available to users' offline processing. It remains to be investigated whether the dominance of SIMILAR IS NEAR over CONSIDERED IS NEAR in the performance data, as well as the dissociation with the subjective judgments, is a permanent (i.e. an effect of general experience) or a temporary effect (i.e. depending on the specifics of the task).

What can be learned from these studies? First, from comparing the results of both experiments, it can be taken that the higher the mental workload, the larger the effects of image-schematic metaphors, because the less likely the cognitive system is to compensate for metaphor-incongruent user interface designs.

Second, the findings in these studies add to the evidence on the psychological reality of the metaphor SIMILAR IS NEAR – DIFFERENT IS FAR. It goes beyond recent findings in the cognitive sciences, since it studies the effects of the metaphor on performance data rather than on similarity judgments alone (cf. Casasanto, 2009b; Montello et al., 2003). Although the results generally complement the findings of earlier studies, no reversion of effects was found between perceptual (pointer displays) and conceptual (number displays) processing, as in the Casasanto (2009b) study. SIMILAR IS NEAR seems to be equally valid for both types of display. Further work should identify whether this is due to the differing complexity of stimuli. In Casasanto's study, the stimuli were quite complex (faces and abstract words) and the similarity of these stimuli could be determined along several dimensions. In the experiments reported here, simple stimuli were used and similarity was determined on one dimension only (i.e. numerical or angular difference). It is therefore necessary to replicate the findings with more complex concepts of similarity (e.g. feature-based similarity, Tversky, 1977).

Third, the findings also shed new light on the proximity compatibility principle (Wickens & Carswell, 1995) that is reflected in the metaphor CONSIDERED IS NEAR – NOT CONSIDERED IS FAR. The results indicate that it sometimes can be more beneficial to put user interface elements far from one another, even when they need to be attended to during the same task. However, as speculated above, whether the proximity compatibility principle is valid or not might depend on the task and user goals.

Summarising the findings of experiments 3 and 4 and considering the caveats discussed above, the following tentative guideline can be derived:

> If you want to represent similarity or difference information by using the spatial dimension NEAR-FAR and performance is important, then use the mapping SIMILAR IS NEAR – DIFFERENT IS FAR. For other tasks and if satisfaction is more important, use the proximity compatibility principle, i.e. the mapping CONSIDERED IS NEAR – NOT CONSIDERED IS FAR.

6.5 Conclusion from Studies 1 to 4

Previous evidence in the cognitive sciences has shown that image schemas and their metaphorical extensions are not just a matter of language but also reflect the underlying structures of thought. In studies 1 to 4 the claims of image-schema theory were investigated in the context of user interfaces. Metaphorical extensions of the image schemas UP-DOWN, LEFT-RIGHT, and NEAR-FAR were used and it was found that cognitive metaphors can be beneficial for designing intuitive use.

Metaphor-congruent interfaces can be used with higher effectiveness, higher mental efficiency, and higher satisfaction than metaphor-incongruent user interfaces. As the studies on the NEAR-FAR image schema have shown, these effects are more pronounced when the task is more mentally demanding. Effectiveness gains of up to 100% (in study 4) and mental efficiency gains of up to 14% point to the importance of user interfaces being congruent with image-schematic metaphors. However, further studies are needed that investigate the effects of the context, the task, and the interaction of such metaphors in user interfaces.

In their results these experiments add substance to two promises of image-schema theory: (1) that image schemas and their metaphoric extensions contribute to intuitive use and (2) that image schemas contribute toward meeting the demand for interacting with abstract information by means of physical/spatial displays and controls. Metaphorical extensions of image schemas were shown to be involved in online and offline processing when interacting with user interfaces, as shown by the performance and satisfaction data, respectively.

The four studies also showed how image-schema theory could extend and complement current guidelines in user interface design. Image-schema theory can provide the theoretical background to the notion of population stereotypes and can contribute a wide range of new stereotypes that until now were not documented in the ergonomic literature. Design principles such as the proximity compatibility principle can be rephrased as metaphorical extensions of image schemas. Such reformulations of current guidelines can help to better understand the kinds of contexts in which they can be applied and whether alternative metaphors can account for contradicting findings.

As we are far from fully understanding the general validity and the detailed effects of image-schematic metaphors, more research needs to follow. The present studies have led to a number of questions to focus on. In addition, a large range of about 250 metaphorical extensions awaits their application and investigation in user interfaces.

7 Inter-Rater Reliability

One of the questions posed in chapter 4 is whether the set of image schemas gathered from the literature can constitute a usable vocabulary of a design metalanguage. The main concern was that image schemas were too interrelated for designers to agree whether a certain aspect of the user's mental model, the task, the user interface, or the interaction with technology represents instantiations of a particular image schema or not. In the following two studies, the inter-rater reliability of the image-schema vocabulary as a prerequisite for using image schemas in user interface design is tested.

7.1 Study 5: FORCE image schemas

For this study the group of FORCE image schemas was chosen, because FORCE image schemas seem harder to extract than image schemas from, for example, the SPACE and CONTAINMENT groups that present themselves more readily to the eye of a user interface analyst. That something is UP or DOWN, IN or OUT, in the CENTRE or in the PERIPHERY is obvious and agreement between raters should be high. SPACE and CONTAINMENT image schemas are often instantiated by static components of the user interface. FORCE image schemas like COMPULSION, MOMENTUM, and DIVERSION, in contrast, are instantiated by the more transient dynamics of interacting with technology.

In context-of-use analyses, these dynamic aspects, like clicking on a button or moving a lever, are often documented as verbal descriptions in static form. This could make FORCE image schemas harder to perceive. Furthermore, while spatial structure is always a concrete property of a user interface, FORCE image schemas can be instantiated in both physical (e.g. blocking the movement of a lever) and abstract ways (e.g. blocking an unauthorised user from accessing a website). In short, the group of FORCE image schemas warrants special attention, because it should be more difficult to categorise than other groups. If the inter-rater agreements within this challenging group of image schemas are acceptable, they probably are for other groups of image schemas as well.

7.1.1 Pre-Study: Generation of Usage Examples

A pre-study was conducted to generate usage episodes that could later be used as material for image-schema categorisations. The study had the sub-goal of exploring whether image-schema definitions can easily be understood and whether people are able to relate them to their own experiences with technology. In a workshop, participants were first introduced to definitions and examples of ten FORCE image schemas. Then, they were asked to brainstorm examples from their experience with user interfaces that matched each of these image schemas.

Method

Participants. Eleven researchers from the Engineering Design and Methodology group at TU Berlin took part in the pre-study. Participants had no prior experience with FORCE image schemas, but they were familiar with the specific method of brainwriting used in the study.

Material and Procedure. The workshop started with 30 minutes of presentation plus 10 minutes of discussion. During the presentation, participants received general information about what image schemas are and how they can be used in user interface design. Then, ten FORCE image schemas were introduced one by one, along with their notations, metaphorical uses, and a discussion of user interface examples.

After the presentation, each participant received one questionnaire sheet. Each sheet stated the name, the notation, and the definition of one FORCE image schema (see appendix A.5 for an overview). Different sheets contained different FORCE image schemas. Participants were instructed to write down examples from their experience with user interfaces that they regarded as instances of the specific image schema on the questionnaire sheet before them. For each example they indicated the direction of the effect, i.e. whether the user influences the technology, or the technology influences the user. They also indicated how confident they were that their example was a proper instance of the specific image schema (coded from 1 = very uncertain to 5 = very certain). To facilitate brainwriting, one example was already printed on the sheet. For instance, the example on the RESTRAINT REMOVAL sheet was "The driver releases the handbrake to move off." For further examples see appendix A.5.1.

Participants were instructed to fill in as many examples as they could think of. After two minutes, they had to pass their sheets to their neighbours on the right. They then immediately started working on the sheets they received from their neighbours on the left. After ten of these two-minute cycles, each participant had had the chance to produce examples for all ten image schemas in sequence.

Results and Discussion

During the 20-minute brainwriting session the participants produced 146 usage examples in total. This averages to 14.6 examples per image schema (*SD* = 3.84) and 13.3 examples per participant (*SD* = N/A). Participant were, on average, quite confident in relating their examples to image schemas (*M* = 3.74, *SD* = 1.19). The highest confidence ratings were given for examples of the image schemas COMPULSION, BLOCKAGE, and ENABLEMENT (*M* = 4.36, 4.15, and 4.13, respectively), the lowest for COUNTERFORCE, DIVERSION, and MOMENTUM (*M* = 3.17, 3.29, and 3.31, respectively). For further details refer to appendix A.5.2.

Summarising the results it seems that 30 minutes of instruction were enough to induce sufficient confidence in the participants to relate their user interface experiences to the descriptions of ten FORCE image schemas. The many examples gained were re-used in the second part of the study in which people familiar with image schemas re-assigned image-schema categories to these examples.

7.1.2 Main Study: Image-Schema Classification

The second part of the study aimed at determining the agreement among people familiar with FORCE image schemas who re-classified the usage examples generated in the pre-study. This procedure allowed estimating the inter-rater reliabilities of image-schema classifications, both among the participants in the main study and as compared to the original classification obtained in the pre-study.

Method

Participants. Four people familiar with FORCE image schemas took part in the study. Two of them (students) had experience in applying the full range of image schemas to the analysis of user interface elements, such as an airbus cockpit (Hahn, 2007) and business software (Weber, 2007). The other two (project collaborators) had each attended two one-day workshops on the application of FORCE image schemas to haptic interaction. They were familiar with the definitions of FORCE image schemas and with generating user interface examples. None of them had participated in or knew about the results of the pre- study.

Material and Procedure. The image-schema examples obtained in the pre-study were reviewed and any examples were removed that were not referring to user-technology interaction at the user-interface level (e.g. *driving a car against a tree*). Also, any duplicate examples and ambiguously phrased examples were removed (e.g. a phrase like *measuring devices* that does not point to a specific interaction episode). To further reduce the amount of examples, only those were included in the study that at least had received a confidence score equal or above three (of a maximum of five). This left 80 examples, which were then slightly edited to complete unfinished sentences, define technical terms like *brake power assist unit,* and add some context information like consequences of the described interaction. Care was taken to not give cues away easily as to which image-schema categories apply to the examples. A sentence like *The user blocked the system by doing XY* would be re-formulated to *The user did XY to prevent the system from….*

This list of examples was given to the participants, together with a 'cheat sheet' containing the ten image-schema definitions, notations, and user interface examples (cf. Table A.41 in appendix A.5.1). Participants were instructed to assign one image schema to each of the 80 examples in the questionnaire. On a five-point scale they indicated for each image-schema assignment how confident they were about their choice (coded from 1 = very uncertain to 5 = very certain).

Different image-schema categories had different prevalence in the list of examples. The categories of ATTRACTION/REPULSION (12 usage examples) and

ENABLEMENT (11 examples) had the highest prevalence. The category of COUNTERFORCE (3 examples) had the lowest prevalence. The other categories had medium prevalence: BLOCKAGE (9), RESTRAINT REMOVAL (9), COMPULSION (9), DIVERSION (8), MOMENTUM (7), BALANCE (6), and RESISTANCE (6).

Statistics. Inter-rater agreement is indicated by Cohen's kappa values (Cohen, 1960; Eugenio & Glass, 2004). In contrast to raw percentage values, kappa takes into account that a proportion of the agreement can occur purely by chance. Kappa values can vary between -1 (complete disagreement) to +1 (complete agreement). A kappa value of 0 indicates chance agreement. Inter-rater agreement can also be computed for single rating categories (here, single FORCE image schemas) and is then called intra-class agreement.

The interpretation of kappa values follows the guidance provided by Landis and Koch (1977). A kappa value of $\kappa < .00$ indicates poor agreement, $.00 \leq \kappa \leq .20$ slight agreement, $.21 \leq \kappa \leq .40$ fair agreement, $.41 \leq \kappa \leq .60$ moderate agreement, $.61 \leq \kappa \leq .80$ substantial agreement, and $.81 \leq \kappa \leq 1.00$ almost perfect agreement.

Note that these guidelines should be interpreted with care as the number and prevalence of categories and rater bias affect the magnitude of the value. The kappa will be lower when there are more categories, when prevalence is not homogenous across categories, and when biases between raters occur (Sim & Wright, 2005). As a consequence, kappa values that are prevalence-adjusted and bias-adjusted (*PAK*) are also reported (for the calculation of *PAK*-values see Sim & Wright, 2005).

Results

Agreement Between Raters. Four participants classified the 80 usage scenarios into ten image-schema categories (see appendix A.5.3 for the agreement-disagreement matrix). The overall kappa value is $\kappa = .59$ (Table 7.1). Using the Landis and Koch (1977) criterion, this is interpreted as a moderate agreement between participants. The kappa values of single image-schema categories are 'almost perfect' in two cases, 'substantial' in two cases, 'moderate' in four cases, and 'fair' in two cases. 'Almost perfect' agreement was obtained for the image schemas ATTRACTION/REPULSION and BALANCE. Only 'fair' agreement was found for the image schemas RESISTANCE and COUNTERFORCE.

Adjusting kappa values for prevalence and bias (*PAK*) had only small effects on most of the ratings, except for COUNTERFORCE, RESISTANCE, and ENABLEMENT. After adjustment, agreement values were much higher for RESISTANCE and COUNTERFORCE, now indicating moderate and substantial agreement. Agreement in the category ENABLEMENT, however, decreased to moderate agreement.

Table 7.1. *Inter-Rater Reliabilities for* FORCE *Image Schemas*

	κ	*PAK*
Overall	.59	.59
ATTRACTION/REPULSION	.95	.90
BALANCE	.81	.81
DIVERSION	.68	.71
ENABLEMENT	.61	.50
MOMENTUM	.52	.59
BLOCKAGE	.52	.48
RESTRAINT REMOVAL	.45	.49
COMPULSION	.45	.48
COUNTERFORCE	.34	.63
RESISTANCE	.30	.51

Agreement with the Standard. This measure indicates how strong the agreement of single raters or groups of raters is with the original classifications made in the pre-study. The overall kappa value is κ = .71, indicating a 'substantial' agreement. The kappa values of single image-schema categories are 'almost perfect' in two cases, 'substantial' in six cases, 'moderate' in one case, and 'fair' in one case (Table 7.2, second column). Again, 'almost perfect' agreement was obtained for the image schemas ATTRACTION/REPULSION and BALANCE. Only 'fair' agreement was found for the image schema COUNTERFORCE. When using prevalence adjustments the value for COUNTERFORCE changed to a 'substantial agreement'. None of the other values changed as much.

Can the agreement with the standard be enhanced? For example, by grouping participants and comparing their combined ratings to the standard? Out of four participants, six possible pairs and four possible triples were formed. Group scores were determined by chosing the image-schema classification that the majority of group members agreed on. Conflicts were resolved by considering the confidence ratings obtained for each classification. If, for example, in a pair of participants one participant classified a usage example as an instance of COUNTERFORCE with a confidence rating of 4 (out of 5) and the other as an instance of BLOCKAGE with a confidence rating of 2, then the classification of the pair was assumed to be COUNTERFORCE and this result was compared with the standard. This procedure served as a simple model for possible negotiation of classifications in a group of designers.

The results show that increasing the number of people classifying examples into image-schema categories enhances agreement with the standard (Table 7.2). In other words, the more raters participate in the classification of image schemas, the less errors occur. When using pairs of raters, the overall agreement with the standard rises to κ = .81, an 'almost perfect agreement' compared to the 'substantial agreement' using single raters. Adding a third and a fourth rater further increases the agreement with the standard.

Table 7.2. *Agreement of Differently Sized Groups of Raters with the Standard Classification*

	single raters		rater pairs		rater triples		four raters	
	κ	*PAK*	κ	*PAK*	κ	*PAK*	κ	*PAK*
Overall	.71	.72	.81	.82	.83	.83	.87	.88
ATTRACTION/REP.	.95	.91	.98	.97	.97	.96	.95	.93
BALANCE	.89	.90	.97	.97	1.00	1.00	1.00	1.00
ENABLEMENT	.76	.67	.92	.89	.89	.85	.95	.93
MOMENTUM	.72	.75	.88	.89	.88	.89	.93	.93
DIVERSION	.67	.71	.74	.77	.72	.76	.78	.80
BLOCKAGE	.67	.64	.72	.71	.79	.78	.86	.86
RESTR. REMOVAL	.66	.66	.75	.74	.80	.78	.84	.80
RESISTANCE	.61	.70	.68	.77	.85	.89	1.00	1.00
COMPULSION	.56	.58	.69	.68	.64	.65	.70	.70
COUNTERFORCE	.40	.72	.57	.83	.56	.84	.65	.86

The kappa values of single image schemas show a similar development. With pairs of raters classifying usage examples, four image-schema categories have 'almost perfect agreement', five show 'substantial' agreement, and one 'moderate' agreement with the standard. It is again the image schema COUNTERFORCE that lags behind, while ATTRACTION, BALANCE, ENABLEMENT, and MOMENTUM are in the top group. A group of four raters delivers even better results: seven categories show 'almost perfect agreement', the other three show 'substantial agreement'. The RESISTANCE category benefitted the most from pooling classifications in a group; the kappa increased from κ = .61 with single raters to a maximum of κ = 1.00 when a group of four raters was compared against the standard. Again, COUNTERFORCE showed the largest difference when using prevalence adjustments (cf. *PAK*-values in Table 7.2).

Qualitative Analysis of Category Confusions. A qualitative analysis of the examples reveals likely sources of category confusions and indicates what can be done to enhance the reliability of image-schema categorisations. The results show that inter-category confusions occur in a range of typical situations:

- when two image-schema instances occur together, especially
  - when one image schema describes a direct consequence of another image schema, e.g. ENABLEMENT as a consequence of RESTRAINT REMOVAL;
  - when one image schema is the precondition of another, e.g. BLOCKAGE is the precondition for RESTRAINT REMOVAL;
- when the direction of the interaction is not specified, i.e. when it is unclear whether the user or the technology should be regarded as the agonist in the interaction;
- when coders deal with non-typical examples of an image schema, e.g. instances of active (instead of passive) ENABLEMENT;
- when coders have not enough information to distinguish similar image-schema categories, e.g. when trying to distinguish COMPULSION from MOMENTUM, information is needed about the presence (COMPULSION) or absence (MOMENTUM) of external forces;
- when coders shift their focus from human-computer interaction to instances of human-human interaction or system-internal interactions.

Knowing these potential causes of confusion, measures can be taken to increase inter-rater agreement (see below).

7.1.3 Summary and Conclusion

Study 5 investigated the inter-rater reliabilities of FORCE image schemas. In the pre-study eleven people brainstormed examples from their daily interaction with technology for each of ten FORCE image schemas. The results showed that people can relate their experiences easily to the image-schema definitions provided and that they did so with high confidence.

These examples were given to four people who were more experienced in dealing with FORCE image schemas. Their task was to assign each of 80 examples of usage scenarios to one out of ten FORCE image schemas they found most suitable. The overall inter-rater reliability was moderate with an overall kappa value of $\kappa = .59$. Kappa values for single image-schema categories ranged between 'fair' and 'almost perfect', i.e. between $\kappa = .30$ and $\kappa = .95$.

Agreement with the originally intended image-schema classifications was higher than the pure inter-rater reliabilities. Single raters achieved an average kappa value of $\kappa = .71$, indicating substantial agreement with the standard. Using more than one rater increased the agreement with the standard up to $\kappa = .87$ when using a group of four raters (indicating an 'almost perfect' agreement).

The practical recommendation derived from these data could be this: if you strive for maximum agreement with a set standard, use as many raters as possible. Good results, however, can already be achieved with two raters. This makes studies that involve image-schema extractions economically feasible in practice.

However, a moderate inter-rater reliability of $\kappa = .59$ still leaves room for improvement. The qualitative analysis leads to the conclusion that a more systematic training of image-schema coders could be beneficial. The training should include the basics of image-schema theory and their application to user interface design, as well as definitions and examples of image schemas. Special emphasis when coding usage scenarios should be put on:

- the subtleties of different image-schema definitions, e.g. of the distinction between passive and active ENABLEMENT;
- the correct identification of the direction of interaction: is the image schema applied to an episode of data input or output?
- the distinctions between causes, results, and further consequences of usage events;
- what information is needed and where this information can be found, e.g. to distinguish BLOCKAGE, RESISTANCE and COUNTERFORCE.

However, because these recommendations are only based on the FORCE group of image schemas, it is necessary to include other groups of image schemas in studying the inter-rater reliability of an image-schema vocabulary.

7.2 Study 6: Image Schemas Extracted from a Context-of-Use Analysis

The last study focused on the one group of FORCE image schemas. They were selected, because it was assumed that their dynamic character and their double occurrence in concrete and abstract instantiations make them less obvious to the analyst than, for example, SPACE and CONTAINMENT image schemas. Whether this is the case can only be determined if FORCE image schemas are compared to other groups of image schemas.

Study 5 was limited to coding pre-fabricated usage examples. They were designed in a way that each example matched one image-schema category. In practice, however, more than one image schema could be instantiated in the same object of analysis, e.g. a user interface widget or a sentence uttered by a user. Because this ambiguity may complicate coding under real-world conditions, inter-rater agreement can be expected to be lower.

The aim of study 6, therefore, was to investigate the agreement between two raters under the conditions of a real-world design project using a large range of image-schema categories. The material from which image schemas were extracted came from a context-of-use analysis of an invoice-verification and posting process conducted in the accounting department of a beverage company (see chapter 8, study 7). Image schemas were extracted from four sources: task steps, users' utterances, the user interface of a software application, and the steps in the user-system interaction.

Another objective was to investigate the 'natural' prevalence of image schemas in the different components of the context of use. It is expected that image-schema instances from the FORCE and PROCESS groups have high prevalence in the dynamic task steps and the user-system interaction. Image schemas from the SPACE, CONTAINMENT, and ATTRIBUTE groups should be most frequent in the static user-interface screens. For users' utterances an even distribution of image-schema instances across groups is expected.

7.2.1 Method

Participants

Two raters participated in the study. The first had a theoretical background in image-schema theory, but applied image schemas in this study for the first time. The second was the author of this work.

Procedure

The first rater conducted a context-of-use analysis by observing and interviewing three users at their workplace (Weber, 2007). She summarised the results of the context-of-use analysis in a table containing a 38-step task sequence, the corresponding user-system interaction, and the users' utterances. Table 7.3 shows the entries in the first row of the table (i.e. the first step in the task sequence); the full table is documented in Weber (2007). In addition, 15 screenshots of the SAP R/3 software application were documented along with brief descriptions of their functionality (see Weber, 2007).

The second rater was not present during the observations at the users' site. He received some global information about the work process from the first rater and was walked through the sequence of screens of the SAP R/3 software. The table and the screenshots provided the basis for image-schema categorisation. Both raters independently assigned image-schema categories to the information provided in the table and the screenshots. The same agreement statistics as in study 5 were computed (kappa-values and adjusted *PAK*-values).

Table 7.3. *Extract from the Documentation of Task Steps, User-System Interaction, and Users' Utterances (Translated from Weber, 2007)*

Task Step	User-System Interaction		Users' Utterances
	User	System	
#1 Starting the invoice verification and posting procedure: • starting the SAP R/3 system • opening an invoice		Shows the screen *Entry* in left display	"Thus, here: Office, Entrance, Workflow, Inbox".
	Clicks on *Workflow*		„Here I learn that I must do something". [if rows are marked in yellow, denoting a passing deadline]
		Opens screen *Inbox* in left display	"These are our yellow slips; these are people whom we've sent something, but who haven't answered yet".
	Marks checkbox in the row of the invoice and clicks on the icon *Execute*		"Now I simply grab one of them". [when opening an object in the inbox] "Then I open this one here".
		Opens the screen *Invoice Verification* in left display and *Image of original invoice* in right display.	"…at the same time the image of the invoice appears here". [when opening an invoice, an image of the original invoice appears on the second display] "…and enter the invoice"

7.2.2 Results and Discussion

Prevalence

In study 5, each object of analysis had been assigned to only one image schema. In this study, an object of analysis can have multiple aspects to which more than one image schema may be assigned. Consider this utterance of a user: "Then one has to compare everything: delivery date, invoice date, invoice number; see whether this is all correct, whether the account number is correct." *Then* denotes following a PATH step by step. The verb *has to* points to an instance of the COMPULSION image schema. *Compare* and *see whether [...] is correct* point to instances of the MATCHING image schema. The different numbers that have to be compared form a COLLECTION. It happened that raters missed one or the other aspect of an object of analysis – whether the object of analysis was a user utterance, a task step, an episode of human–computer interaction, or a user interface element.

Thus, the overall number of extracted image-schema instances is different between the raters. Summarised over all context-of-use components (task, interaction, utterances, and screens), the first rater extracted 605 instances of image schemas; the second rater extracted 552. Of these two sets, 434 image-schema instances referred to the same aspects. This equals 72% of the first raters' and 79% of the second raters' number of aspects rated. These numbers show that one rater alone would miss a substantial number of aspects that other raters regard as important. A practical consequence would be using teams of raters, to be able to capture most aspects of the object of analysis.

Note that agreement between raters does not tell whether these image schemas are valid or not. Aspects that were not assigned to image schemas therefore cannot be categorised as misses or true rejections.

A similar phenomenon has been discussed in detecting usability problems (Hertzum & Jacobsen, 2003). The average agreement between any two evaluators of the same system ranged from 5% to 65% ($M = 22.4\%$, $SD = 19.8$) over a range of 12 studies employing different techniques (cognitive walkthrough, heuristic evaluation, thinking aloud). Because of these low detection rates, it is usually recommended to increase the number of testers needed. Compared with these numbers, the agreement between raters in the present study is quite satisfying.

How are image-schema categorisations distributed over image-schema groups? The CONTAINMENT group received the most categorisations (Table 7.4, *Overall* column). Second was the FORCE group, followed by MULTIPLICITY and SPACE. The least frequent categorisations occurred in the group of PROCESS image schemas. Note that the BASIC image schemas did not receive categorisations on common objects, and that the ATTRIBUTE group had only one categorisation (image schema BIG-SMALL). Due to their extremely low prevalence, both groups were omitted from further analyses.

Looking at the single components of the context of use, the predictions about the prevalence of image-schema categories can be tested. As predicted, the prevalence of image-schema groups is different across the four components of the context of use, $X^2(12) = 56.75$, $p < .001$. In the task steps, the groups of FORCE and MULTIPLICITY received the highest number of assignments. This result is probably a consequence from the specifics of the task: users had to complete many steps of comparing data sets with other data sets (e.g. invoices versus orders, digitised data versus original data) that resulted in MULTIPLICITY categorisations (e.g. MATCHING, LINK). Missing data, missing and granted authorisations, as well as verifying and correcting data led to a large number of FORCE image schemas (e.g. BLOCKAGE, ENABLEMENT, DIVERSION). The number of PROCESS image schemas was not enhanced remarkably.

Table 7.4. *Prevalence of Image-schema Groups Extracted from Different Context-of-Use Components*

Image-schema Group	Overall	Task steps	User-System Interaction	Users' Utterances	User Interface
CONTAINMENT	37%	17%	30%	36%	55%
FORCE	26%	31%	36%	28%	11%
MULTIPLICITY	19%	32%	21%	13%	18%
SPACE	17%	18%	8%	22%	15%
PROCESS	2%	2%	5%	2%	0%

Because the user-system interaction is dynamic, a large proportion of FORCE image schemas was expected and becomes apparent in the data (Table 7.4). The generally less frequent PROCESS image schemas were most frequent here. Quite astonishingly, a high proportion of CONTAINMENT image schemas were categorised – mainly due to frequent FULL-EMPTY classifications based on users checking whether certain data were available on the screens or not. The SPACE image schemas in their turn were not very frequently used in describing the human-computer-dialogue.

User's utterances featured mainly CONTAINMENT and FORCE image schemas as well. There was also a large amount of SPACE image schemas. Although not evenly distributed (as had been predicted), the distribution of image schemas across groups most closely match the distribution that can be expected from the *Overall* column in Table 7.4.

Finally, the user interface screens featured many instances of CONTAINMENT (buttons, windows, group boxes, icons) – more than half of all categorisations. This left less room for other categories. The SPACE image schemas had also been expected to be very frequent. This was not so much the case. It could be that other image-schema instances were more salient than instances of SPACE image schemas. The raters often omitted rating the position of user interface elements, i.e. whether UI elements were in the LEFT or RIGHT, in the CENTRE or the PERIPHERY. Similarly, ATTRIBUTE image schemas either did not apply (e.g. HEAVY-LIGHT, STRONG-WEAK) or were not salient enough to be noticed by raters (e.g. BIG-SMALL, DARK-BRIGHT).

In sum, some general patterns emerged – like a high number of SPACE and CONTAINMENT image schemas in the user interface screens, and a higher number of FORCE image schemas in the more dynamic task and interaction steps. Other category assignments, like the high frequency of MULTIPLICITY in the task steps, are probably specific to the invoice-verification task of the users. In conclusion, different patterns of outcomes have to be expected from different sources of a context-of-use analysis. Further research is needed do discern what patterns can be expected generally across different tasks and what patterns are task-dependent.

Agreement

Regarding aspects categorised by both raters, the overall agreement was ‘substantial’ (Landis & Koch, 1977) with $\kappa = .66$ ($PAK = .68$). If image schemas are not analysed as single instances, but as members of image-schema groups, the kappa value is similar, $\kappa = .69$ ($PAK = .72$). This indicates that most confusion takes place between, and not within, image-schema groups (if it were the other way around, a larger change in kappa values could be expected).

Substantial agreement was obtained for the groups of CONTAINMENT, FORCE, and SPACE image schemas (Table 7.5, *overall* column). Moderate agreement was achieved in the groups of MULTIPLICITY and PROCESS image schemas, although for the latter with an excellent prevalence-adjusted kappa. The BASIC and ATTRIBUTE groups of image schemas were not included in the analyses because of their small prevalence.

The single components of the context of use vary widely in their inter-rater reliabilities (Table 7.5). Overall agreements were fair *(κ)* to moderate (*PAK*) for the task steps, and moderate for the user-system interaction (last row in Table 7.5). The agreement on the user’s utterances was substantial, and almost perfect on the user interface screens.

Clearly, the reliabilities on the task steps and the user-system interaction are not satisfying. Is there an explanation for the moderate performance? Although both raters assigned the image schemas based on the same documentation, they may still have worked on different information. The first rater had first-hand experience in interviewing and observing the users. The second rater had not as much experience with the context of use and based his decisions on a short introduction to the context of use and the documentation. A qualitative analysis revealed that the first rater often considered links and common themes between task steps, while the second rater was more tied to the description of single steps, hardly considering the consequences of earlier steps. In addition, the task steps and the user-system interaction were documented very sparsely. In contrast, the users’ utterances and the user interfaces gave a more complete basis for analysis. The result was a greater agreement in these categories.

A practical consequence of these differences is to include as many analysts in the user observations as possible. If this is not feasible, analysts not taking part in the observation should be provided with a richer information basis. A more comprehensive experience could be created, for instance, by letting analysts retrace the users’ work steps themselves using real data from the context-of-use analysis, or even let them access and use the same technology the users apply.

Table 7.5. *Inter-Rater and Intra-Class Agreement for Different Context-of-Use Components*

Image-schema Group	Overall	Task steps	User-System Interaction	Users' Utterances	User Interface
CONTAINMENT	.78 (.69)	.52 (.52)	.38 (.30)	.87 (.81)	.98 (.97)
FORCE	.72 (.67)	.19 (.07)	.88 (.78)	.76 (.71)	.95 (.95)
SPACE	.67 (.72)	.49 (.48)	.37 (.65)	.66 (.67)	1.00 (1.00)
MULTIPLICITY	.57 (.62)	.49 (.35)	.11 (.20)	.70 (.79)	.94 (.97)
PROCESS	.44 (.92)	1.00 (1.00)	.23 (.73)	.49 (.92)	--
Overall (image-schema groups)	.69 (.72)	.41 (.42)	.46 (.49)	.75 (.77)	.97 (.97)
Overall (single image schemas)	.66 (.68)	.40 (.42)	.45 (.48)	.68 (.69)	.94 (.95)

Note. The first number in each cell denotes the value of kappa (κ), the number in brackets denotes the prevalence-and-bias adjusted kappa (*PAK*).

Rating image schemas in groups instead of single image-schema categories enhanced agreement rates only slightly (last two rows in Table 7.5). Again this shows that most disagreement stems from confusing groups of image schemas rather than image schemas within groups.

And what can be said about the intra-class reliabilities of image-schema groups? The hypothesis was that FORCE image schemas should be more difficult to categorise than SPACE or CONTAINMENT image schemas and thus should lead to lower agreement between raters. Indeed, study 5 showed that the inter-rater agreement was only moderate for FORCE image schemas. The present study could not confirm this hypothesis, though. Overall, the FORCE and SPACE image schemas had similar reliabilities. The intra-class reliabilities varied across context-of-use components. The SPACE and CONTAINMENT image schemas had higher intra-class reliabilities in rating the user-interface screens and tasks steps. The FORCE image schemas fared better in rating the user-system interaction and the users' utterances. The agreement in the MULTIPLICITY group peaked for users' utterances and user interface screens. Agreement rates for the PROCESS group are difficult to judge, because they were influenced by their low prevalence rates (as seen from the large differences between κ and *PAK* values).

Looking at the reliability rates that imply less than substantial agreement (i.e. all $\kappa < .61$), what recommendations can be given to enhance them? The high agreement rates for users' utterances and user interfaces point out that raters found sufficient clues to base their categorisations on. On a user interface, for example, everything with a boundary denotes a CONTAINER. In users' utterances, verbs can be associated with FORCE image schemas and prepositions with SPACE and CONTAINMENT image schemas. Such mappings are less clear (or less established) with task steps and interaction episodes. One way of improving agreement rates would be to establish mapping rules that everyone agrees on.

The two raters already used their own mapping rules. As these rules differed between raters, stable patterns of disagreement emerged. One rater, for example, consistently categorised the systems' opening of a new screen as an instance of SPLITTING, the other consistently categorised it as instances of CONTAINER and IN-OUT. The first rater had described her interpretation of the user's view, i.e. data belonging together in one work step are SPLIT among several screens. The second rater had referred to what the new screen looked like (like a CONTAINER) and what the user did (shifting the attention INTO this CONTAINER).

There are many recurring events in the user's task for which common rules may be established. Among them are users searching for something, data exceeding limits, users opening screens, users contacting other people, the user following if-then selection rules, users comparing something, or users authorising something. The goals of further research need to be: (a) to discover recurring patterns in human activity relevant to human-technology interaction and (b) to validate the image schemas that are appropriate to describe them. The goal would be a dictionary of rules for enhancing the reliability of image-schema coding.

The example above shows that agreeing on rules can become complicated, as it is difficult to find the 'right' image-schematic interpretation. To guarantee the validity of these rules, i.e. that the image schemas categorised are actually the ones that are effective in different contexts of use, is certainly the harder task for further research.

7.3 Summary and Conclusion

Image schemas seem to form a vocabulary that is easily understood at first encounter. People easily relate their experience to definitions of FORCE image schemas and can provide usage examples for a range of image schemas (study 5). When experts assign image schemas to these usage examples, they moderately agree on which image schemas to assign. They agree substantially or almost perfectly with the originally intended image-schema classifications on these usage examples.

Extending the scope to a larger range of image schemas and to a real-world context of use revealed that, overall, raters substantially agree on the image schemas assigned (study 6). Agreement rates, however, largely depend on the context-of-use component the image schemas are extracted from. It was found that agreement rates are almost perfect regarding user interfaces, are substantial for users' utterances, but are only moderate for task steps and instances of the human-computer dialogue.

The studies in this chapter yield a range of recommendations for improving the reliability of image-schema categorisations:

- Provide enough information. If information is sparse as in the one-sentence usage scenarios in study 5 or the task and interaction descriptions in study 6, then reliabilities are only moderate.
- Use more than one rater. In study 6 it was found that only about three quarters of the aspects rated by one rater are also rated by the other. Using more than one rater will enhance the overall number of image schemas detected. Study 5 showed that using pairs of raters is best when agreement with a standard needs to be enhanced.
- Provide rules. Raters will make up their own rules in rating similar usage situations. If raters applied common rules, large amounts of disagreement could be removed (study 6).
- Check the validity of the rules. Agreed-upon rules can enhance the inter-rater agreement. But it does not follow that they are also valid in a user-interface design context. Rules need to be empirically tested in different contexts of use.
- Train the analysts. Although even novices can apply image schemas directly, raters could still profit from a systematic training. The training should include the presentation of image-schema definitions and rules for assigning them to different components of the context of use, as well as extensive exercise and feedback, especially discussing the sources of disagreement.

Again, it should be noted that reliability is important in establishing image schemas as the vocabulary of a design language. However, checking the validity of assigned image schemas remains necessary. Additional research needs to determine whether and how image-schema instances in user interfaces actually trigger image schemas in the user's mind, i.e. whether they contribute to intuitive use, are neutral, or even hamper it.

Finally, it cannot be the goal for practitioners to achieve complete agreement between raters. As in language, sometimes the most interesting discoveries are made if people disagree on how to name things. Disagreement between raters, if acknowledged, can lead to fruitful discussions and eventually to better, more creative, and better informed design decisions.

Can image schemas be established as a metalanguage for design? As we are only beginning to learn to use the language, the inter-rater reliabilities found in the above studies may be judged to be acceptable. Understanding each other seems possible. The next chapter takes the idea of image schemas as an abstract design vocabulary a step further and investigates the practicability of applying image schemas in a human-centred design process.

8 Image Schemas as a Metalanguage in Design

The first promise of image-schema theory is that, because of their ubiquity in structuring concrete and abstract thought, image schemas can be used as the vocabulary of a metalanguage for user interface design. With the help of image schemas it should be possible to describe the structure of tasks and users' mental models and to prescribe the structure of user interfaces on an abstract level. One goal of the work described in this chapter is to show the feasibility of applying image schemas throughout a human-centred-design lifecycle. A second goal is to explore possible methods and tools that support the designer in applying image schemas. As a result of these explorations, the practicability of image schemas in user interface design is assessed.

8.1 Study 7: Image Schemas Throughout the Design Process

Previous studies have applied image schemas to single phases of the design lifecycle – the analysis of the context of use (Bakker et al., 2009; Raubal, 1997) and the design phase itself (Antle et al., 2008, 2009; Lund, 2003; Lund & Wiberg, 2001). In the last chapter it was shown that image schemas can be used to analyse usage scenarios (study 5), users' utterances, task steps, user interfaces, and the interaction with the user interface (study 6).

However, the application of image schemas throughout a whole design cycle has rarely been tried so far and the costs and benefits of using the approach are hardly known. Therefore, in the study presented here, all phases of the human-centred design cycle are covered. Image schemas were applied to the re-design of an accounting application. The study involved the development of two different user interface concepts based on the same image-schematic requirements. This was done to test whether image schemas are flexible enough to be applied to different interaction paradigms and to cross-validate the evaluation results. The first concept was a graphical user interface (GUI) that could be employed within a standard personal-computer environment with a mouse, a keyboard, and a display. The second concept combined graphical with tangible user interface (TUI) elements (hybrid TUI-GUI). Tangible user interfaces belong to a developing interaction paradigm that incorporates real-world tangible objects to enrich the bandwidth of human-technology interaction (cf. Ishii, 2008; Ishii & Ullmer, 1997). Using the TUI paradigm in this study allows exploring how usefully image schemas can be applied to the design of non-standard user interfaces.

In the redesign, a user-centred design process according to ISO 13407 was followed comprising the phases (1) understand and specify the context of use, (2) specify user requirements and organisational requirements, (3) produce design solutions, and (4) evaluate designs against requirements (cf. chapter 4.1). Usually, these steps are iterated several times. The first cycle ends with a user-interface concept that is refined in the following cycles until the last cycle ends with the final product. As the goal here is to assess the general feasibility of image schemas in design, the study went only through the first cycle. The result was a user-interface concept that was evaluated with users. As this was not a running prototype, the evaluation focussed on measuring user satisfaction, leaving out performance measures. The methods and results of all phases are described in the following sections.

8.1.1 First Step: Context of Use Analysis

The first activity – understanding and specifying the context of use – comprised the analysis of users, their tasks, the existing technology, and the socio-technical context in which they interact. Methods included user interviews, field observations and reviews of documents, particularly work instructions. The results of this context-of-use analysis were already used as the basis of the inter-rater reliability study in chapter 7.2 (study 6). Here, a more detailed description is given with a focus on the design process.

Method

Observations and interviews followed the methodology of contextual inquiry (Holzblatt, Wendell, & Wood, 2005), according to which observing a small number of users is usually sufficient, because there is only a limited number of ways to perform a task. Three users were interviewed and observed at their workplace in the accounting department of a large German beverage corporation. The users were 42, 51, and 55 years old and had 2, 15, and 17 years of experience with the task. Each user visit lasted about two hours. During observation, users were asked to think aloud and their utterances were recorded using a voice recorder. Additional records about the workflow were made using pen and pencil. Screenshots of the current system were taken as well as photographs of the work environment.

Context-of-use data was processed from four different sources (cf. study 6). First, a task model was built from the observational interviews and the review of work instruction documents. The model located the task of the users within the general workflow in the company and also identified the single steps made by the users. From the task steps in the workflow, image schemas were extracted (cf. Table 7.3). Second, it was assumed that the users' mental model of the task is revealed in their utterances made when thinking aloud. The utterances were transcribed from the audio recordings and image schemas were extracted (cf. Table 7.3). Third, image schemas were extracted from the user interface of the currently used SAP R/3 software. Fourth, image schemas were extracted from the documentation of the users' interactions with the system (cf. study 6).

In addition to observing and interviewing the users, four users filled in the AttrakDiff questionnaire (Hassenzahl, Burmester, & Koller, 2003) about the current SAP R/3 system. AttrakDiff measures how users experience the attractiveness of a product on the subscales *attractiveness* (ATT), *pragmatic quality* (PQ) and *hedonic quality* (HQ). Items of the subscale attractiveness capture the global rating of the perceived quality of the product. The subscale pragmatic quality describes the usability of the product and indicates how successfully users can achieve their goals with the product. The subscale hedonic quality indicates to what extent the functions of the product enhance the possibilities of the users, stimulate them or communicate a particular identity (e.g. by creating a professional, cool, or modern impression). Hedonic quality is divided into the factors *stimulation* (HQ-S), which stands for the novelty, the stimulation, or the creativeness the product conveys, and *identification* (HQ-I), i.e. how well the product helps the user to connect with other people, whether it is stylish or precious. The response format is a seven-point semantic differential with values ranging from -3 to +3.

Results: Users, Tasks, and Technology

A typical user workplace in the accounting department is shown in *Figure 8.1*. Each user has two computer displays. One display contains the SAP R/3 input masks. The other display contains scanned images of invoices, a list of contact persons, and a software tool for taking notes and forwarding them to colleagues.

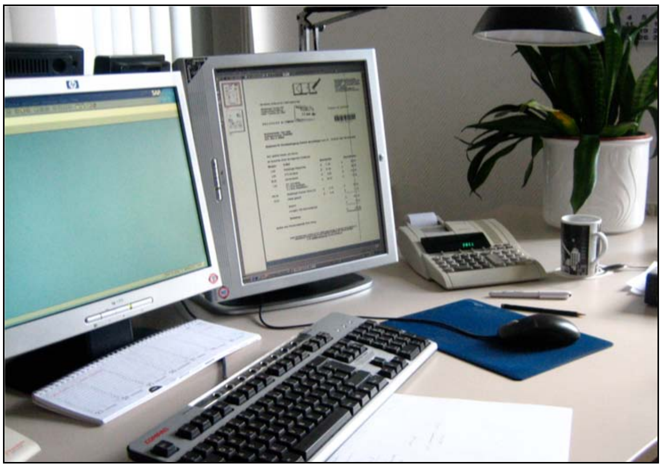


Figure 8.1. Typical workplace with two screens.

The task of invoice verification and posting is not overly complex. The enterprise purchases goods or services and receives an invoice from the supplier in return. Invoices are collected at a central location and are scanned and processed by an optical character recognition (OCR) system. Before the invoice is paid, it is verified in form and content (for an overview see *Figure 8.2*).

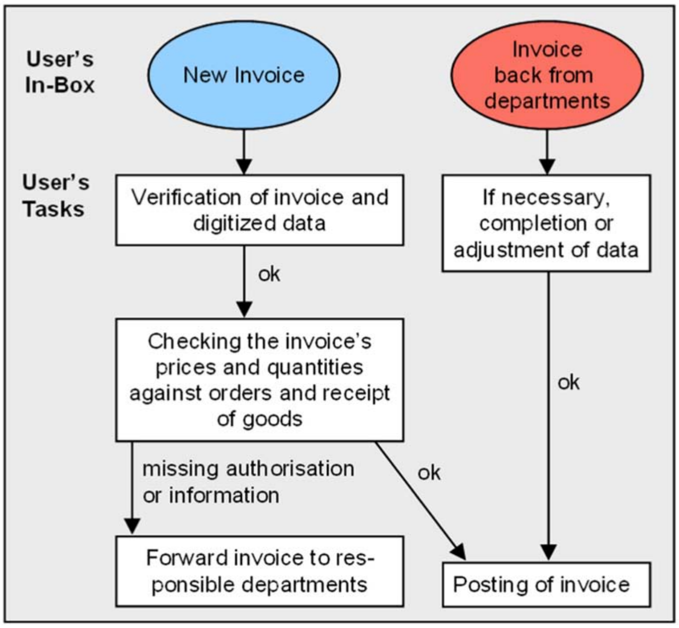


Figure 8.2. Simplified model of the invoice verification and posting process

Users each process about 50 to 80 invoices per day. Invoices usually arrive in the users' inboxes that are part of the SAP system. After opening an invoice, the users check whether the original invoice contains the legally required details, e.g. tax rate, tax number, date of delivery, and discounts given. Then they compare the digitised data on the left display with the invoice image on the right display to make sure everything was recognised correctly in the OCR process. In the next step, prices and quantities of goods received are checked against the original orders. In many cases the orders and the invoice do not match, because orders are not delivered in one lot, or unexpected shipping costs were added. The users also check whether the receipt of goods has been documented in their system. Once everything is complete, the invoice can be posted. Sometimes information is missing or unclear – e.g. no orders are connected to the invoice (approximately in one third of the cases) or the received goods have not been entered into the system. In these cases the users forward the invoice to their colleagues in the operating departments. Their colleagues then provide the data or authorisations needed and send them back to the accounting department. Then, after another check, the users finally post the invoice. If invoices are incomplete or incorrect, the users return them to the supplier and delete them from the system. For a more detailed description of the task see Weber (2007).

The company worked with the software package SAP R/3, version 4.6c. Users rated the software as being average on all four scales of the AttrakDiff questionnaire (using a range of -3 to +3): pragmatic quality ($M = 0.29$, $SD = 0.56$), hedonic quality-stimulation ($M = 0.00$, $SD = 0.39$), hedonic quality-identity ($M = 0.25$, $SD = 0.49$), and attractiveness ($M = 0.18$, $SD = 0.49$).

Being in the neutral range of the scale, however, does by far not mean that the software is flawless. When forwarding the invoices to their colleagues, for example, the users often have to write notes and messages, frequently using prefabricated text fragments. The note editor, which they use for that purpose, is unable to store those text templates. However, users found a clever workaround: they use the email signatures in MS Outlook as templates. They select a preconfigured signature, insert it into an empty email message and then copy and paste it into the note editor. Other flaws of the software included that necessary data was distributed across several screens and that the software required the users to make many selection and confirmation steps for no apparent reasons.

Results: Instances of Image Schemas

Image schemas could be extracted from most aspects of the context of use: task steps, users' utterances, user-system interaction, and user interfaces (cf. study 6 above). Results of the image-schema analyses of the tasks, the user-system interaction, and the users' utterances often complemented each other. To illustrate this, the example of opening invoices from the inbox is described here in more detail to enable the reader to follow one example through all steps of the design cycle. For details of other task steps see Weber (2007).

The actual task of invoice verification and posting starts with opening an invoice from the inbox. This step is linked to the image schemas IN-OUT for the process of opening and the CONTAINER image schema represented by the inbox itself. The step also contains the COMPULSION image schema, as the user's action compels the otherwise idle system to open an item. Data from the users' utterances supports these results. Users said, for instance: *Now I grab one of them* or *Then I open this one here* which again represents the COMPULSION and IN-OUT image schemas. In the case of *grabbing* items from the inbox, the CONTACT image schema can be added to the list.

After an invoice is opened, the user checks whether the data of the paper invoice was digitised correctly. In order to do so, users compare the digitised data with the original invoice image. Because the two different sets of data have to fit together, this step is linked to the image schema MATCHING. The users' language supports this image-schema category: *Then one just has to compare everything, […] to see if everything matches.*

User interfaces can be easily described using image schemas. The instances of image schemas found in the UI can be either congruent with or contrasting the image schemas extracted from the task or the users' utterances. Consider the example of the inbox in *Figure 8.3*. The screen on the left appears after the users have logged in to the system. It is intended to be used for selecting the invoices shown in the inbox. It shows a COLLECTION of buttons (CONTAINERS) arranged one below the other (UP-DOWN). Some of the buttons are nested in a group box (CONTAINER) that itself is nested in the blue area (CONTAINER) of the screen (CONTAINER).

The visual relation between the *Total* button and the group box appears like a PART-WHOLE relation. *Total* implies that all invoices are displayed after pressing this button. The other buttons appear to call up subsets of invoices that, if the numbers beside the buttons were summed up, equal the total sum of invoices. This PART-WHOLE appearance, however, is misleading – the sum of parts (35) is more than the perceived whole (24). In fact, the buttons within the group-box container represent PARTS and WHOLES themselves, e.g. *Office* (WHOLE) and *Office ungelesen* (German for *Office unread*, PART).

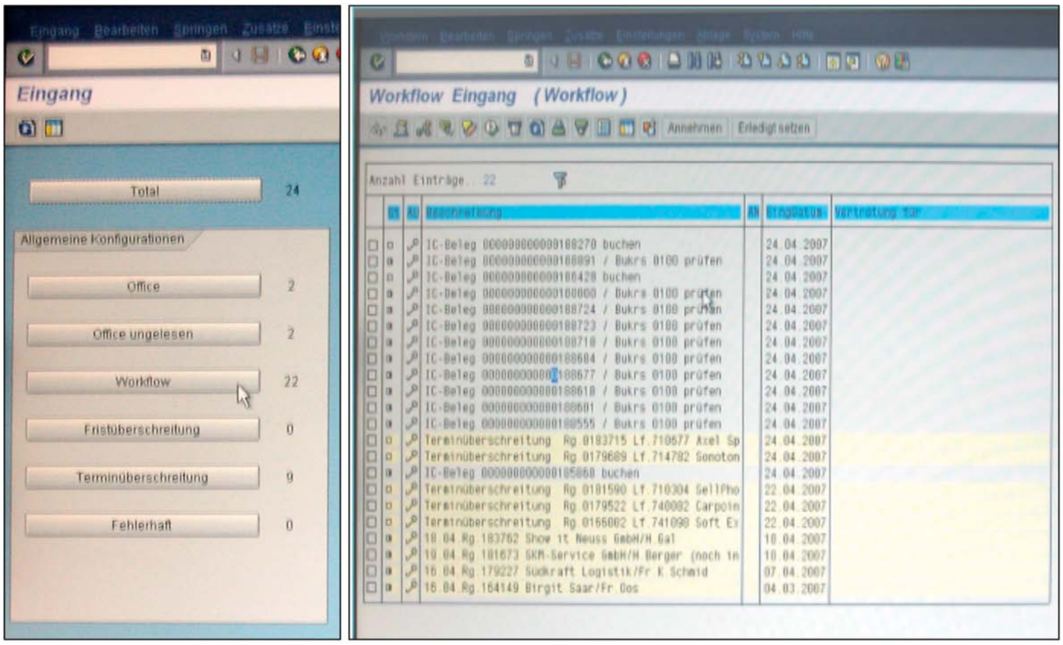


Figure 8.3. Inbox selection screen (left) and inbox with invoices (right)

When the user clicks on one of the buttons (image schema COMPULSION), the actual inbox screen appears (right side of *Figure 8.3*). The specified set of invoices that have not been posted yet is shown in a list (COLLECTION, UP-DOWN, and CONTAINER). The list contains new invoices as well as invoices that have been returned from other departments to which they were sent with the request to provide further information or authorisation. Overdue items that need special attention are marked in yellow (ATTRACTION) – a feature that supports usability as evidenced in the users' utterances: *Here I know that I must do something*.

To open an item from the inbox, the user marks the checkbox on the left side of the item (LEFT-RIGHT, FULL-EMPTY) and then presses the icon with the clock-symbol located above the inbox (COMPULSION, UP-DOWN). Note that this method differs from what was found in the users' utterances. They talked about opening invoices as *grabbing* them – involving direct CONTACT and COMPULSION. Lacking fit between the users' mental model and the actual interaction with the system points out areas in which the current user interface can be improved.

Table 8.1. *User and Task Requirements for the Inbox, with Associated Image Schemas*

	Requirements	Image Schemas
I01	Store all items in one inbox.	CONTAINER, COLLECTION
I02	Provide the possibility to sort items in inbox by date.	UP-DOWN
I03	Provide the possibility to filter items in inbox.	PART-WHOLE
I04	Provide the possibility to find certain items in inbox.	MATCHING, CONTAINER
I05	Support meeting payment deadlines.	ATTRACTION
I06	Provide the possibility to open an item from inbox.	IN-OUT, COMPULSION, CONTACT

In this way the whole series of 38 work steps was analysed, relating image schemas from the users' task, the user-system interaction, the users' utterances, and the user interface to each other. For detailed results of this see Weber (2007).

8.1.2 Second Step: Requirements Specification

In the requirements phase, a high-level description is derived of how the user interface should support the users in solving their tasks. The basis for formulating the requirements were those task steps that where indispensable to the business process of invoice verification and posting plus the related image schemas. This, for example, excluded unnecessary steps imposed by the current technology. To these requirements, image schemas that occurred exclusively in the users' utterances were added. Requirements were formulated in a way that avoids prescribing concrete user interface solutions, e.g. group boxes, tables, or radio buttons. Each requirement was accompanied by a number of image schemas that were associated with its content.

Three sets of requirements for the user interface were created, each focussing on a coherent set of sub-tasks of the user: (1) using the inbox, (2) verifying and editing invoices, and (3) forwarding and posting invoices. Table 8.1, for example, shows the list of requirements for the inbox. Note, how the requirement *I06 – Provide the possibility to open an item from the inbox* and the linked image schemas IN-OUT, COMPULSION and CONTACT take both the work steps and the users' language into consideration that were analysed in the previous section.

Similar tables have been created for the invoice verification process and the forwarding of an invoice (details in Weber, 2007). The comparison of invoice data with the digitised image of the invoice resulted, for example, in the requirement *V04 – Support the user in comparing and matching digitised data against the* *original image of the invoice* associated with the image schema MATCHING. Note that, although this requirement seems rather trivial, the current system does not provide any solution for it. The users spend unnecessary time and energy with the often hard-to-read image of the original invoice to find the data they need to compare, e.g. the bank details.

8.1.3 Third Step: Design Solutions

Starting from the requirements, the production of design solutions followed the users' task flow from opening the inbox to the forwarding or posting of invoices. In producing the hybrid TUI-GUI solution, the first step was to allocate functions to either digital GUI elements or physical TUI elements. This was done with the help of so-called PIBA-DIBA lists (Physicality Is Better At – Digitality Is Better At). This approach is inspired by the MABA-MABA lists (Machines Are Better At – Men Are Better At) introduced by Fitts (1951). For a more detailed description of these lists and their application see Hurtienne, Israel, and Weber (2008).

The next step included the development of detailed user interface elements. Each (image-schematic) requirement was independently attended to, and for each image schema several alternative solutions were sketched. For example, when the requirement was to design a CONTAINER for the inbox, in the GUI a sketch could include a window frame or a group box. In the GUI-TUI hybrid solution, a sketch could include a glass bowl or a wooden box to fulfil the same requirement. The design alternatives were put into a morphological box (Zwicky, 1969; see Table 8.2) and suitable solutions were selected. The criteria for selection were (1) how well the single solutions fit together, (2) how well they represented the implications of the context-of-use analysis, and (3) how appropriate they seemed in the sociotechnical context of current office work. The selected alternatives were then synthesised into the overall user interface solution. In informal evaluation steps, designers checked the developed solutions against the requirements and made changes when necessary.

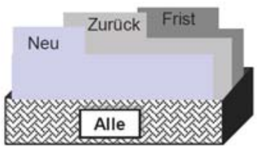
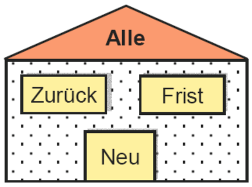
The graphical solution was designed by the same person who conducted the context-of-use analyses and derived the requirements. Two more designers were involved in the design of the hybrid solution: one with an extensive background in image-schema theory, the other a novice with regard to image schemas, but familiar with the design of tangible user interfaces.

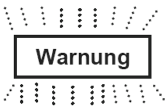
Result I: Design Solution – GUI

For the image schemas connected to each requirement, several alternative GUI instantiations were derived. A morphological box was used to select alternatives and to combine these into the overall user interface solution. Some examples are shown in Table 8.2. The first row, for instance, shows three alternative ways of instantiating the PART-WHOLE image-schema requirement (I03). The second alternative was eventually chosen, because users referred to the initial screen as their *inbox tray*. The tree view in Alternative 1 seemed too abstract, the house metaphor seemed too far removed from the context of use.

Table 8.2. *Morphological Box with Design Solutions for Image-schematic Requirements (Examples)*

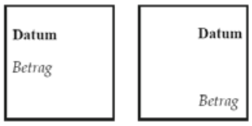
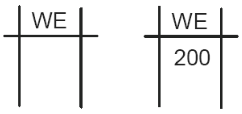
<table>
<tr><td>Image Schema and Requirement</td><td>Alternative1</td><td>Alternative 2</td><td>Alternative 3</td></tr>
<tr><td>I03 – PART-WHOLE
Provide the possibility to filter items in inbox.</td><td>Radio Buttons, arranged hierarchically

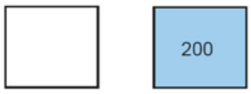
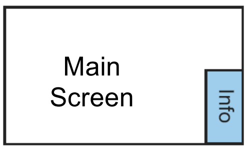
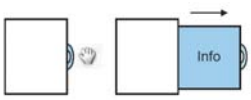
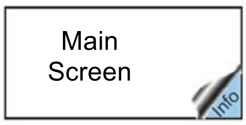
</td><td>Card index box

</td><td>Picture metaphor

</td></tr>
<tr><td>I05 – ATTRACTION
Support meeting payment deadlines.</td><td>Colour

</td><td>Blinking

</td><td>Icon

</td></tr>
<tr><td>V04 – MATCHING
Support the user in the comparison of two datasets</td><td>Same location

</td><td>Same Colour

</td><td>Same Appearance

</td></tr>
<tr><td>V06 – FULL-EMPTY
Display goods-receipt related to invoice.</td><td>With or without entry

</td><td>Picture metaphor

</td><td>Background filled

</td></tr>
<tr><td>F04 – FRONT-BACK
Show auxiliary information.</td><td>Tabbing

</td><td>Sliding

</td><td>Peeling

</td></tr>
</table>

Note. Selected alternatives are marked with a black frame. After Weber (2007, p. 70).

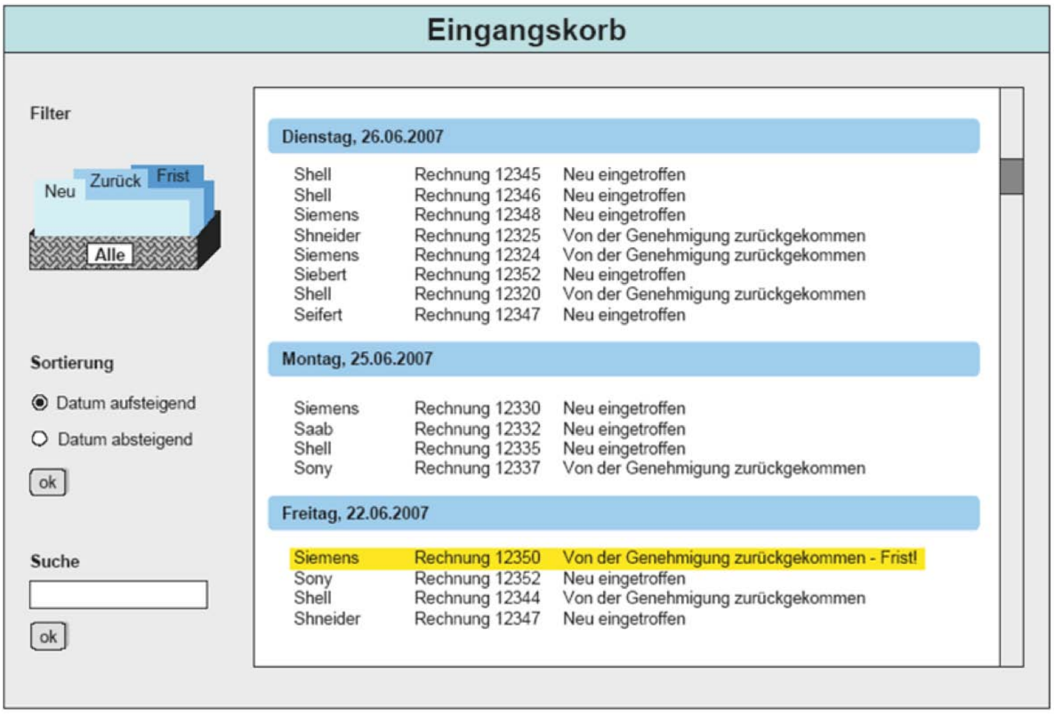


Figure 8.4. The redesigned inbox (cf. Weber, 2007, p. 72)

The redesigned solution consisted of three main screens that were in line with the three sets of requirements mentioned above: using the inbox, verifying and editing invoices, and forwarding invoices.

Figure 8.4 shows the redesigned inbox as an example. Note that all requirements are met by this design solution. Items are COLLECTED in a single CONTAINER (requirement I01), they can be sorted by date (I02, UP-DOWN), filtered (I03, PART-WHOLE), and searched (I04, MATCHING, CONTAINER). The yellow marking for drawing the users' attention to items that are close to deadlines is retained from the current system (cf. *Figure 8.3*). There were other possibilities to design for an ATTRACTION image schema (I05, listed in Table 8.2), but they were felt being either too strong (blinking) or too weak (symbol) to attract the users' attention. Finally, the act of opening an invoice is now much closer to the image schemas found in the users' utterances. Now, they can grab the invoice and open it in a single step – a double-clicking the invoice is sufficient (I06, IN-OUT, COMPULSION, CONTACT).

After opening the invoice, the first task of the user is to compare the original image of the invoice with the digitised data in the system. As in the current solution, both are shown at the same time on different screens. A new feature was introduced to support the comparison task. When users want to find corresponding data on the original invoice, they simply put the mouse pointer over the respective data fields in the new verifying and editing screen (*Figure 8.5*). A coloured highlight that is present in the data screen and in the matching area on the invoice image is cutting down search time. This function is possible, because the system reads the data directly from the digitised invoice. Thus, the exact location of the data is known to the system. In this way, requirement V04 was met following alternative 2 in Table 8.2.

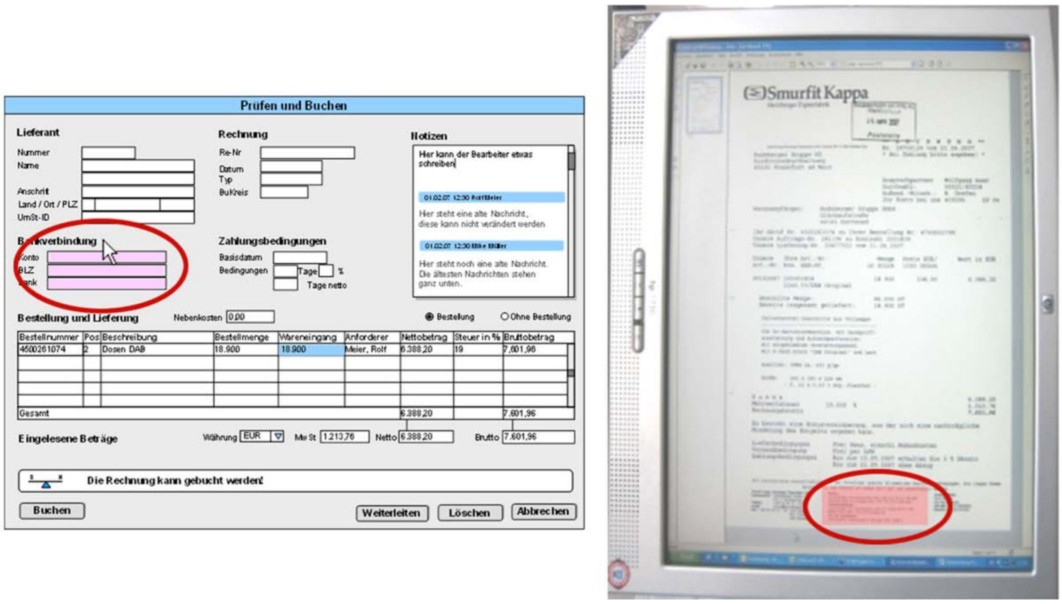
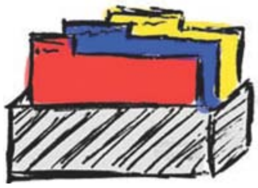


Figure 8.5. Supporting the comparison of data using an instance of the MATCHING image schema (cf. Weber, 2007, p. 76)

Result II: Design Solution – Hybrid TUI-GUI

Function allocation. Using the PIBA-DIBA lists, functions were allocated either to graphical or tangible user interface elements. As the inbox was very object-oriented and the context-of-use data specified simple object manipulations such as grasping, moving, opening and forwarding, it was decided that the inbox should become tangible. For the invoice verification process, however, the graphical user interface was retained, because many different variables were to be represented at the same time, text had to be entered and edited, and so on. A graphical user interface also allows users to access other functions, e.g. e-mail or intranet, which are used for secondary tasks. This, however, did not exclude the use of tangible tokens to augment the user's interaction with the GUI (see below). Similarly, for the forwarding feature, graphical user interface elements were chosen, enabling search (e.g. for contact persons), composing text messages, and so on. However, the final posting and the error check are tangible because these are distinctive actions in the working process.

Table 8.3. *Tangible UI Alternatives for the Inbox*

Image Schema and Requirement	Alternative1	Alternative 2	Alternative 3
I03 – PART-WHOLE Provide the possibility to filter items in inbox.	Card index 	Bowl with differently coloured balls 	Dispenser with compartments 

Note. The selected alternative is marked with a black frame.

Design alternatives. Similar to the design of the GUI solution, several alternative solutions were sketched for each image-schematic requirement for the TUI elements and then the most appropriate alternatives were chosen. As an example, Table 8.3 shows three alternatives to physically represent the PART-WHOLE image schema that is needed for filtering items in the inbox.

Solution overview. The final prototype was a tabletop graphic interactive system with a number of tangible user interface elements (*Figure 8.6*). Two digital displays are embedded in the desk (1), the left one for showing the application data, the other for showing the digitised image of the original invoice. The following tangible elements are included: the inbox bowl with invoice balls (2), text template stamps (3), posting and forwarding mounds (4), and a pen for GUI interaction (5). Two digital search fields are provided with the invoice bowl and the forwarding mound (6).

The inbox bowl provides a CONTAINER for the invoice balls (COLLECTION, requirement I01). The bowl is transparent so that the user can easily determine the level of balls, reflecting the amount of work left to do. Invoice balls arrive at the bottom of the bowl through a flexible tube. The arrival of an invoice is thus audible as well as visible.

Invoices are pre-filtered into new, returned, and urgent invoices (PART-WHOLE, I03). The status of each invoice is indicated by coloured LEDs inside the balls. The mapping between physical balls and digital invoices is automatically set by the system. New invoices are glowing red and returned invoices blue. Urgent invoices have a slowly pulsating yellow light (ATTRACTION, I05). The PART-WHOLE character is emphasised by additionally sorting the differently coloured balls into different sections of the inbox bowl. Colours can be digitally re-assigned, so that new or urgent messages can digitally “bubble” from the bottom to the top of the ball stack without the need of physically re-arranging the balls

(I02, UP-DOWN). For tracking the balls' positions within and outside the bowl, and for switching their status, RFID technology can be employed (cf. Hinske, 2007).

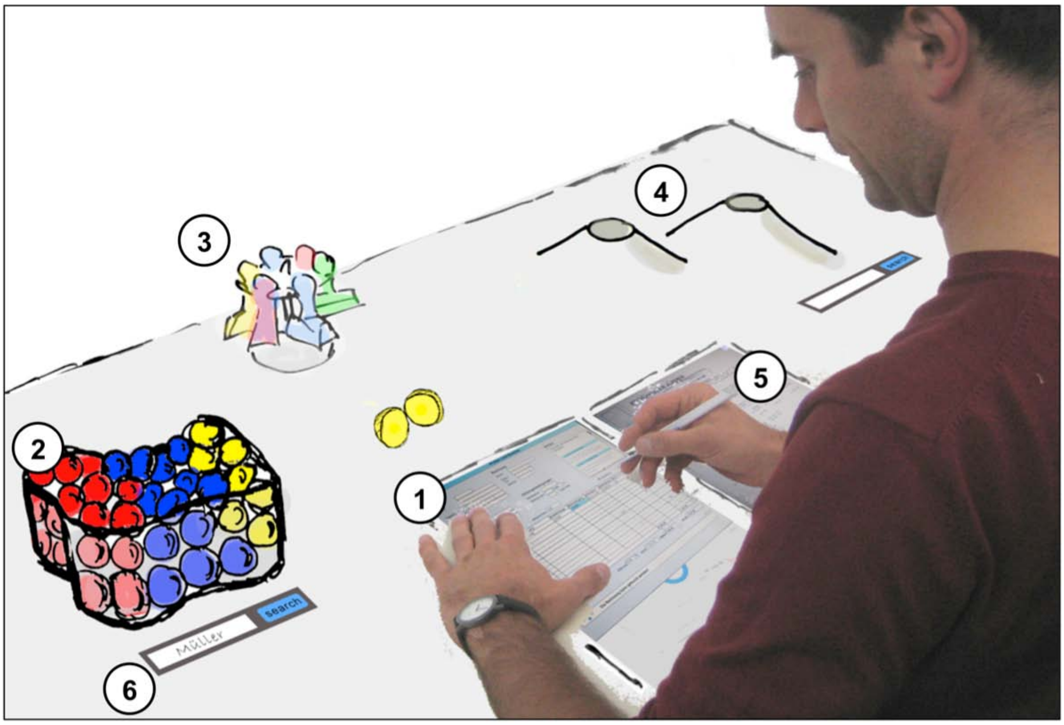


Figure 8.6. Overview of the redesigned business application including TUI and GUI elements

Note that this design reflects the data already available in the current software solution. The inbox used today only marks invoices with meaningless 18 digit numbers, and an additional mark indicating "post" or "verify" (*Figure 8.3*). Urgent invoices (passing their payment deadline) currently have a yellow background.

Specific invoices can be searched with the help of the search field underneath the bowl (CONTAINER). Keywords can be entered with the pen. Then, MATCHING invoices are indicated by blinking and 'bubble' to the top for easy access (I04, UP-DOWN). To start working with an invoice, the user takes a ball out of the bowl (IN-OUT, CONTACT) and opens it to access the invoice's content on the screen (*Figure 8.7* left, COMPULSION, I06).

Other UI elements. Data are edited directly on the screen by using a pen and electronic handwriting recognition. The screen for invoice verification and posting was taken from the GUI solution. Often, for notes and messages, standard text templates are used which are stored in tangible stamps (*Figure 8.6*). If one of these stamps is pressed onto the note editor in the GUI, the template text associated with the stamp is digitally transferred.

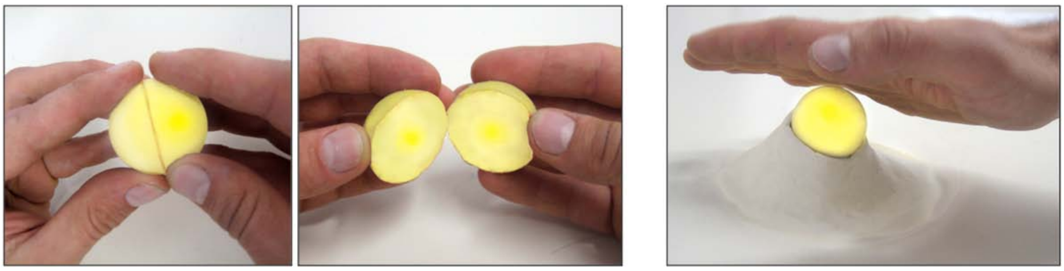


Figure 8.7. Manipulation of the invoice balls. Left: Opening an urgent invoice ball. Right: Posting it by gently pressing the invoice ball into the opening of the posting mound.

Once the verification and editing of an invoice is finished, the user closes the invoice ball and puts it into either the posting mound for posting or the forwarding mound for forwarding. Both mounds require the user to instantiate a COMPULSION image schema, because the balls have to be gently pressed into the openings of the mounds to finally achieve forwarding or posting (*Figure 8.7*, right). If the forwarding or posting cannot take place, because of errors in the invoice, the ball will pop back out of the mound (BLOCKAGE, ATTRACTION).

8.1.4 Fourth Step: User Evaluation

After the two user interface concepts were finished, five users (including the interviewed users) took part in an evaluation session. In a half-day workshop, a simplified version of the pluralistic usability walkthrough technique (Bias, 1994) was employed. All users gathered around a large table. Then, the two designers presented their understanding of the current invoice-verification and posting process using a storyboard on paper. This was used as a warm-up, to provide a common ground on how the workflow is understood. It also served to validate the designers' assumptions about the process.

Then the re-designed applications were presented in the same way – step by step using paper storyboards. Users were asked to comment on the prototypes and voice their opinions. First, the GUI solution was presented, then the hybrid TUI-GUI solution. After each walkthrough, users filled in the AttrakDiff questionnaire for the new prototypes. Additionally, users rated nine statements that compared the re-designed prototype with the current SAP R/3 system on a number of issues, among them whether users thought they would work faster with it or whether the new solution would support the users' work better (see Table 8.4 below). The answer scale reached from 1 (does not apply at all) to 5 (applies completely). The differences between solutions were statistically analysed using the Fisher-Pitman randomisation test for small samples (Bortz & Lienert, 1998).

Qualitative Results

Qualitative results – GUI. During the pluralistic walkthrough users gave much feedback regarding single features of the redesigned user interface. They liked the new solution better, because it saved them many working steps compared to the current SAP R/3 system. With the new solution they would not need to search for orders and goods-receipts because they are shown in the main screen. The MATCHING function that supports users in comparing the invoice data (*Figure 8.5*) was seen as favourable. It was judged as 'comforting the eye' because most of the data on the original invoice is in small print or otherwise hard to read (e.g. because of stamps printed across it). They also liked other features, like the inclusion of the note editor in the main screen, which they found more comfortable to work with. Users also provided the designers with further comments on the details of some features, e.g. where certain system messages should appear.

Qualitative results – Hybrid TUI-GUI. The users liked the hybrid solution very much, because it opened up new possibilities on how to interact with a computer and how to go about accounting work. They said they found the solution *cute* and *different from the usual grey-in-grey* of current computer user interfaces. They imagined themselves passing each other invoice balls instead of forwarding invoices through the software system. Instead of using invoice balls as containers for invoices, they suggested customising them to the beverage company, i.e. using small beer bottles that then could be opened with a bottle opener. Because the transparent inbox with the invoice balls let everyone see how much work is left for the day, they required a second inbox with fewer balls that they could substitute for the original in case the boss is paying a visit. They thought using the prototype would be fun, but were unsure whether it was appropriate to have fun working in an accounting department.

Users were unsure whether handling physical balls would increase or diminish health problems like symptoms of repetitive strain injury (RSI). They were sceptical about some practical aspects: wouldn't the balls wear off at the hinges? Would they add extra cost for maintenance and administration? Would the system be able to recognise peculiar types of handwriting and might the recognition rate not be too slow? Would using the template stamps disturb colleagues by making too much noise?

Quantitative Results

Figure 8.8 shows the comparison between the AttrakDiff ratings of the existing SAP R/3 system and the newly designed solutions (see Appendix A.6 for details). However, all results should be regarded with care, as these are only based on a small number of participants (N = 5). The redesigned GUI solution was rated higher in pragmatic quality than the SAP R/3 system, Fisher-Pitman randomisation test $p < .05$. The analysis of the single items showed that users found the graphical solution less technical and less complicated than the SAP R/3 system. Users rated the hedonic quality 'stimulation' marginally higher, $p < .10$, but not the hedonic quality 'identity', ns. Users rated the attractiveness marginally higher for the redesigned GUI solution than for the SAP R/3 solution, $p < .10$. The analysis of the single items showed that they found the redesigned GUI solution better overall, more likeable, and more pleasant than the SAP R/3 solution.

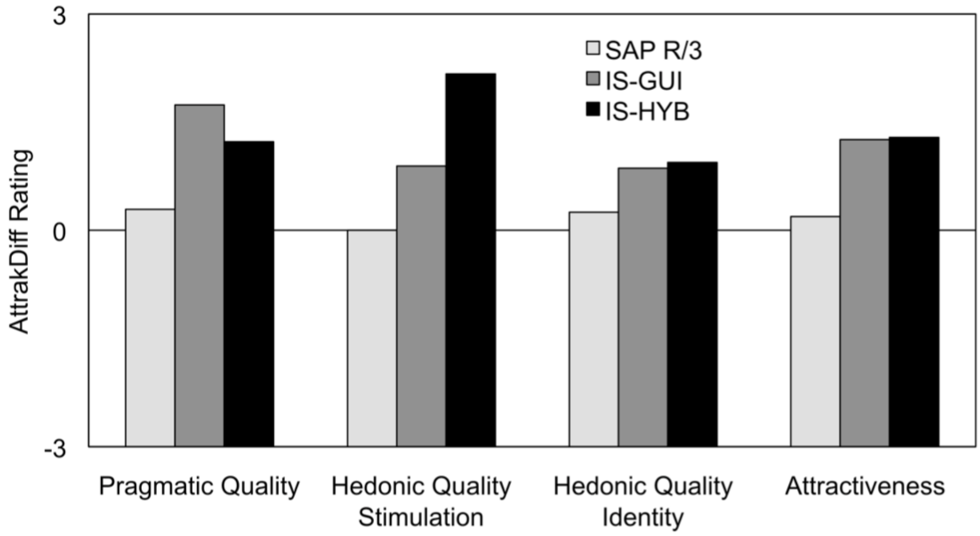


Figure 8.8. AttrakDiff results of the SAP R/3 system, the graphical prototype (IS-GUI), and the hybrid tangible-graphical prototype (IS-HYB)

The hybrid TUI-GUI solution was significantly higher in attractiveness than the original SAP R/3 solution, Fisher-Pitman randomisation test, $p < .05$. The analysis of the single items showed that the users rated the hybrid solution to be more likeable and pleasant than the SAP R/3 solution. The improvement in pragmatic quality was marginally significant, $p < .10$. Users rated the hybrid solution to be less technical, less complicated, and better manageable than the SAP R/3 solution. The largest improvement was made in the hedonic quality 'stimulation', $p < .01$. Users rated the hybrid solution more inventive, bolder, more innovative, more captivating, and more novel than the SAP R/3 system. There were no significant improvements in the hedonic quality 'identity'.

The hybrid solution did not differ from the redesigned GUI solution except in the hedonic quality 'stimulation' in which it was far better, $p < .05$. Users rated the hybrid solution to be more inventive, bolder, more innovative, and more novel than the redesigned GUI system.

The answers to the additional questions are shown in Table 8.4. Comparing the redesigned GUI and hybrid solutions, the hybrid solution was rated less favourable than the GUI solution. The GUI solution was rated positively in all items (i.e. all $M > 3.5$) while the hybrid solution was mainly rated in the neutral range (i.e. $2.5 < M < 3.5$), except for one item . The detailed analysis shows that users were more likely to imagine themselves working with the GUI solution than with the hybrid solution, $p < .10$ – also over extended periods of time, $p < .10$.

Users also indicated that they would have better work support ($p < .10$), would work faster ($p < .01$), and would be more satisfied ($p < .10$) with the GUI solution than with the hybrid solution. All other differences between the GUI and the hybrid solution were not significant.

Table 8.4. *Results of the Additional Questions Regarding the Redesigned Solutions*

	IS-GUI		IS-HYB		Sign.
Item	*M*	*SD*	*M*	*SD*	*P*
I can imagine working with it. [m]	4.20	0.84	3.20	0.84	.095
I can imagine working with it over extended periods of time. [m]	4.20	0.84	3.20	0.84	.095
This solution supports my work better than the current solution. [m]	4.20	0.84	3.00	0.71	.052
I would work faster with this solution than with the current solution. **	4.40	0.55	2.25	0.50	.008
I would be more satisfied working with this solution than with the current solution. [m]	4.20	0.84	2.75	0.96	.056
Using this solution I training new colleagues would be better than with the current solution.	4.00	1.00	3.00	1.22	.151
This solution better corresponds to the image I have of my tasks than the current solution.	3.80	1.64	2.50	1.00	.151
If it were possible I would recommend buying this solution to my company.	3.80	1.64	2.75	1.71	.238
I would recommend implementing this solution to the SAP AG.	3.80	1.64	2.50	1.73	.175

Note. IS-GUI is the image-schematic graphical user interface prototype. IS-HYB is the image-schematic hybrid tangible-graphic user interface prototype. The answer scale ranged from 1 (does not apply at all) to 5 (applies completely). The significance *P* refers to the differences between the prototypes (Fisher-Pitman randomisation test, ** $p < .01$, ^{m} $p < .10$).

In summary, regarding the hybrid prototype, AttrakDiff values were above average, but the answers to the more practical questions ranked lower than those given for the GUI redesign. Thus, the ambiguity of the users between fascination about the innovative approach and scepticism about practical matters that were found in the qualitative results is mirrored in the quantitative data. Perhaps, the large difference between using the current SAP R/3 system and the relative unknown way of working with the hybrid solution is too big a change in working to be readily accepted by the users (cf. Berner, 2006).

8.1.5 Discussion

This study had two goals – to assess the feasibility of image schemas throughout a human-centred design cycle and to make an early assessment of the particular costs and benefits involved.

Feasibility

In this study, image schemas were applied to all phases of a user-interface design cycle. It was shown that during the context-of use analysis image schemas could be extracted from single task steps, the users' interaction with the system, and the utterances of the users (for inter-rater reliabilities see study 6 above). From the results of the context-of-use analysis, requirements for the user interface were formulated. Image schemas were associated with the requirements to provide structural constraints for design.

The selection of the image schemas to be included in the requirements list was determined by the designer and needed careful consideration. The PATH image schema, for example, that was found in the task sequences, was not included in the requirements for the user interface. Because users spend their whole working time at this task, they know the sequence of task steps very well. Instantiating these steps in a PATH image schema would not have added value for the users. The decision to leave out PATH instantiations was approved by the users in the evaluation session. To avoid arbitrary subjective decisions on what to include in the requirements, common sense and user involvement are necessary. In more complex projects, user participation needs to take place very early in the design cycle. Users should evaluate the requirements list before the project enters the design phase, because the designers' decisions may be biased and late changes are costly (cf. Hurtienne, Prümper, & Rötting, 2009).

The context-of-use analyses also showed that image schemas can be useful in the evaluation of user interfaces. Usability problems with the SAP R/3 system were identified where the image-schematic structure in the user interface differed from the image schemas found in the task and users' utterances.

When applying image schemas as a metalanguage, it is expected that image-schematic requirements can be used to prescribe the structure of design solutions. Study 7 has shown that image schemas can guide the production of design solutions without overly constraining the creativity of the designer. Using the method of the morphological box (Table 8.2), many different instantiations of image schemas can be derived - from the serious to the playful. The study also showed that image-schematic requirements can be instantiated equally well in graphical and tangible user interface elements. Image schemas provide the skeletal structure of what to achieve in the user interface and leave much room for the creativity of the designer to fill in the concrete instantiations according to context.

The evaluation (section 8.1.4) has shown that users rated the redesigned prototypes better in pragmatic quality and attractiveness than their current system (which was rated to be neutral). Users preferred the redesigned GUI solution over the current system with regard to speed of use, work support, training, and satisfaction. Users preferred the hybrid solution over the current system with regard to hedonic quality, pragmatic quality, and attractiveness. The hybrid solution even surpassed the GUI prototype in hedonic quality-stimulation. At the same time, users evaluated the hybrid solution more critically with regard to the practical aspects of use.

Costs and Benefits

While study 7 provided evidence that image schemas can be successfully applied to a user interface design process and that with their help viable user interface prototypes can be designed in principle, other questions still remain open. What costs are added by involving image schemas in the process and what benefits can be expected? The results of the study lead to several speculations.

First, ahead of any project there is the question: how much knowledge and training is needed for user interface designers to use image schemas? The answer is: probably very little. The list of image schemas provided in chapter 1 is immediately applicable as has been shown in various contexts. The designers involved in study 7 had either none or very little prior experience in applying image schemas. A formal training would be recommended to refine the sensitivity to and knowledge of image schemas (not least to enhance reliability), but it is not necessary to *start* applying image schemas in design.

Second, for the planning phase: It seems to be reasonable to integrate image schemas into all steps of the design process to receive the full benefit of applying them. Using image schemas during context-of-use analysis and requirements specification does not add much value unless they are also used in the phases of design and evaluation. Using image schemas in design and evaluation is only possible if it is clear what the image-schematic requirements are.

Third, the largest chunk of costs in using image schemas in the design cycle occurs during the context-of-use analysis. It is assumed that as in a normal design process, the analysis of the task, the users' utterances, the users' interaction with the system, and the screens of the existing system have already taken place. To extract image schemas from these sources would require an extra of about 15%-20% of the time needed for the traditional context-of-use analysis. There are possible ways of reducing this effort to help in reducing overall costs. For instance, the analysis of the screens of the current user interface could be reduced to only the most frequently used screens. It should also be explored whether image schema extraction from the users' utterances can be done without first transcribing the utterances, because this can be very time consuming, if done manually. Not much extra effort applies in the requirements phase – the most salient image schemas from the context-of-use are associated with the user interface requirements – and those need to be specified anyway.

Fourth, the investment into image-schema analysis pays off in the design phase. Once image schemas are formulated as requirements, they easily guide design decisions without being overly restrictive. Image schemas contribute to reducing the ‘design gap’ between the requirements and design phases. They advise the designers what to implement in the user interface in a more concrete way than requirements without image schemas. They also have the advantage of splitting the design task into smaller chunks. Designers then brainstorm instantiations of single image schemas and later synthesise them into an overall solution. Using this procedure, a number of interesting and creative design alternatives emerge, as illustrated in Table 8.2. Image schemas introduce an abstract level of thinking about design; thus freeing the designer from sticking to the concrete form of how things were done in previously designed solutions. Image schemas help in streamlining the design process. Using them as a metalanguage can save the time otherwise needed for re-describing or re-thinking requirements before producing user-interface designs. This can be particularly important in projects where the requirements analysts are not the same people as the user interface designers.

Fifth, in evaluation, image schemas provide a vocabulary for describing design problems occurring in user interfaces. They also hint at possible solutions. This was shown in the example of the misleading PART-WHOLE image schema in the inbox selection screen above (*Figure 8.3*). Again, employing image schemas can save time and effort, because the problem, the required solution, and the design can be described in a single language.

As said above, much of this is still theory. So far, there is only subjective and qualitative evidence on the usefulness of image schemas to a design process, while the quantitative estimates of costs and benefits still have to be calculated. Further research needs to compare a traditional UCD process and an image-schema augmented UCD process. From the results, it could be determined more precisely whether and how much value image schemas add, measured e.g. in terms of number of design solutions produced, quality of the final product, or time saved in the design process. It will be essential that further studies include high-fidelity (i.e. running) prototypes. With the low-fidelity prototypes used in the previous study only one component of intuitive use was measured: the satisfaction of users, expressed in their subjective ratings. High-fidelity prototypes will allow the evaluation of the interaction’s effectiveness and mental efficiency, too.

8.1.6 Conclusion

The exploration of image schemas in design practice shows that there is reason to believe that they have practical value. Image schemas are readily extracted from the results of a context-of-use analysis and can be directly transferred to the requirements list. They can help analysts to make explicit what is implicit in the users’ task and mental model and describe it using a fairly abstract vocabulary. This helps practitioners to focus on the essential aspects of the user interface and to avoid using preconfigured design solutions. Using image schemas can positively distract designers from finding solutions that merely centre on the deployed technology. Image schemas are abstract enough to catch the essentials of what is required of the user interface without dictating design solutions, and they can be applied to designing tangible and graphical user interfaces alike.

Image schemas do not add much cost but promise many benefits for a user-centred design process. However, a direct comparison of design processes conducted with versus without image schemas is needed to determine the cost and value added by image schemas.

Apart from measurement issues, some pragmatic issues need to be considered. In their abstractness, image schemas can only provide the skeleton of a future UI design. Designers may appreciate this, because it provides guidance without restraining creativity. Using the morphological box as design tool has helped to focus on producing design solutions for single image schemas before synthesising them into the overall design concept. But this strength of image schemas could also be their weakness. Providing great flexibility is good – but only those designers will profit that have a certain level of creativity, of experience with building interactive systems, and of proficiency at user-friendly design.

Can a better support be provided to inspire designer's creativity, to provide ideas for image-schema instances across different kinds of user interfaces, and to help establish a better general understanding of image schemas in user interface design? These questions are addressed by the development of a database of image schemas in user interfaces – described in the following.

8.2 Supporting Design with Image Schemas: The ISCAT Database

From the results of the previous studies, several requirements were formulated that could make a database useful in different phases of a user-centred design process:

- Supporting the phase of context-of-use analysis. To enhance the reliability of image-schema extractions, analysts need a good command of image-schema definitions and their interrelations. As image-schema definitions sometimes make subtle distinctions (e.g. the distinction between active and passive ENABLEMENT or the object and observer view of NEAR-FAR), analysts need a possibility to consult definitions of image schemas. Currently, descriptions of image schemas are scattered across the literature. Hence, the database should provide a central repository of definitions and descriptions of image schemas, their relations to other image schemas and typical user tasks.
- Supporting the phases of producing design solutions and evaluation. Using requirements lists including image schemas, designers may wish to get inspired by image-schema instantiations in current user interfaces. To support this, the database needs to hold (and make available) a range of user interface

examples from a variety of interactive products, along with evaluations of the effects these image-schema instances have and possible constraints of their use.

- Supporting the phase of producing design solutions. Even without employing image schemas during the whole design process, designers might want to look up how user interface elements can be designed to convey certain abstract meanings. A question might be: "How can I convey the idea of *importance* to the user?" As this relates to metaphorical extensions, designers would benefit from a centralised collection of metaphorical extensions already known that are currently scattered throughout the literature.
- Supporting the extension and customisation of the database. The research community should be able to add to the database other and new instances of image schemas in user interfaces. Furthermore, versions of the database could be created that are individually tailored to companies using it and that could hold company-specific instantiations of image schemas.

Taking these requirements into account, a database was built during the research for this work that provides a rich set of information on image schemas and their applications in user interface design. The database is realised as a Zope / Plone content management system with a custom add-on written in Python.

8.2.1 ISCAT Overview

The user interface of ISCAT (Image-Schema CATalogue) is structured in two main parts. The first part contains image-schema descriptions and metaphorical extensions found in language (*Figure 8.9*). Each image-schema description contains a definition, its entailments, and synonyms. It contains notes on the experiential grounding of the image schema and lists relations to other image schemas. Attached to the entries of image schemas are their metaphorical extensions as found in the literature (Baldauf, 1997; Jäkel, 2003; Lakoff et al., 1991; Lakoff & Johnson, 1980, 1999; Stefanowitsch, 2006; and many more). Metaphorical extensions are documented with their target domains, experiential grounding, and linguistic examples from different languages (not shown in *Figure 8.9*). At the time of writing, the database contains 45 image schemas and 255 metaphorical extensions. Each description of an image schema is linked to image-schema instances from user interface studies (right hand column in *Figure 8.9*).

The second part, user interface studies, contains a number of user interface examples. A user interface study can be divided into sub-studies and user interfaces. For instance, *Figure 8.10*, shows a flap lever (user interface) in the sub-study *Centre Pedestal* of the study *Cockpit A320.* Each user interface contains one or several images depicting the user interface element and a short description of what the UI element is for and what it does. All image-schema instances that were extracted from the UI element are shown below this information. They include a description of why the analyst thought that the image-schema category applies to the interface, the effect that is achieved by the image-schema instance, how strong the effect is, whether it supports, hinders or is neutral to usability (indicated by a smiley icon), and whether the analyst thought this is a prototypical instantiation of the image schema (indicated by the light-bulb icon). Further notes can be given regarding the decisions the analyst had made in analysing an instance of the image schema.

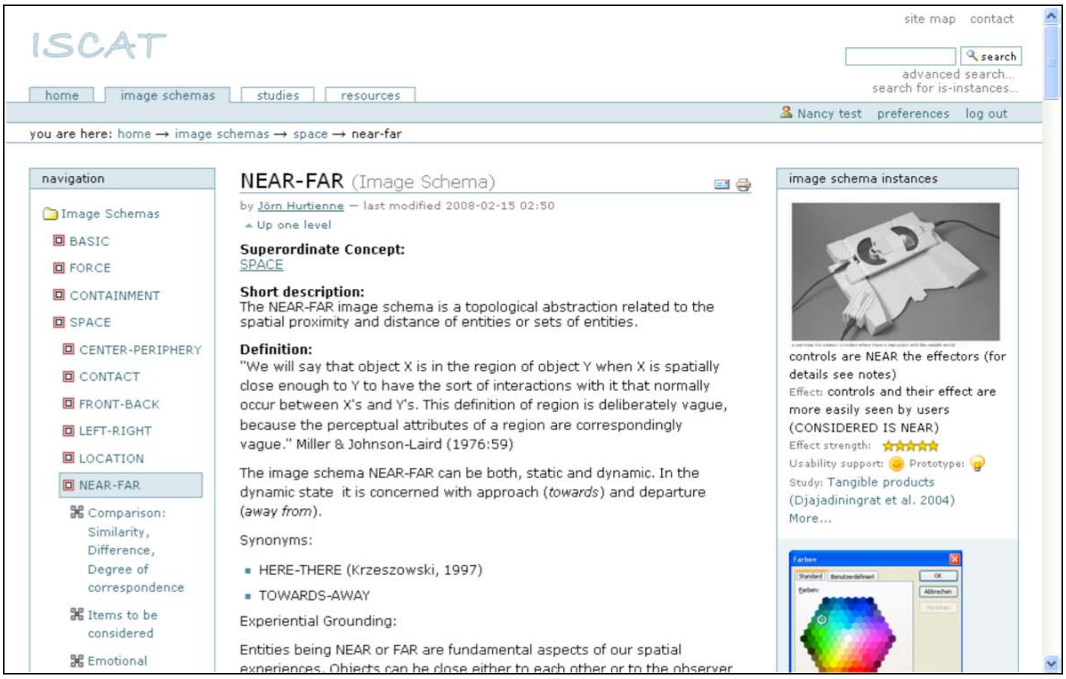


Figure 8.9. Example screen from the ISCAT database, page view of the image schema NEAR-FAR

User interface examples were collected by extracting image-schema instances from cash and ticket machines (Dong, 2007), enterprise resource planning systems (Weber, 2007), and an Airbus 320 cockpit (Hahn, 2007). Further, image-schematic analyses of documented user interface elements were conducted by the author and included in the database, e.g. from the Microsoft Windows styleguide (Microsoft Press, 1995), descriptions of tangible user interfaces (Djajadiningrat et al., 2004; Hurtienne & Israel, 2007), and web interfaces (Fitzpatrick, 2001). At the time of writing, the database contains 30 studies and sub-studies, 197 user interfaces, and 622 instances of image schemas in user interface elements.

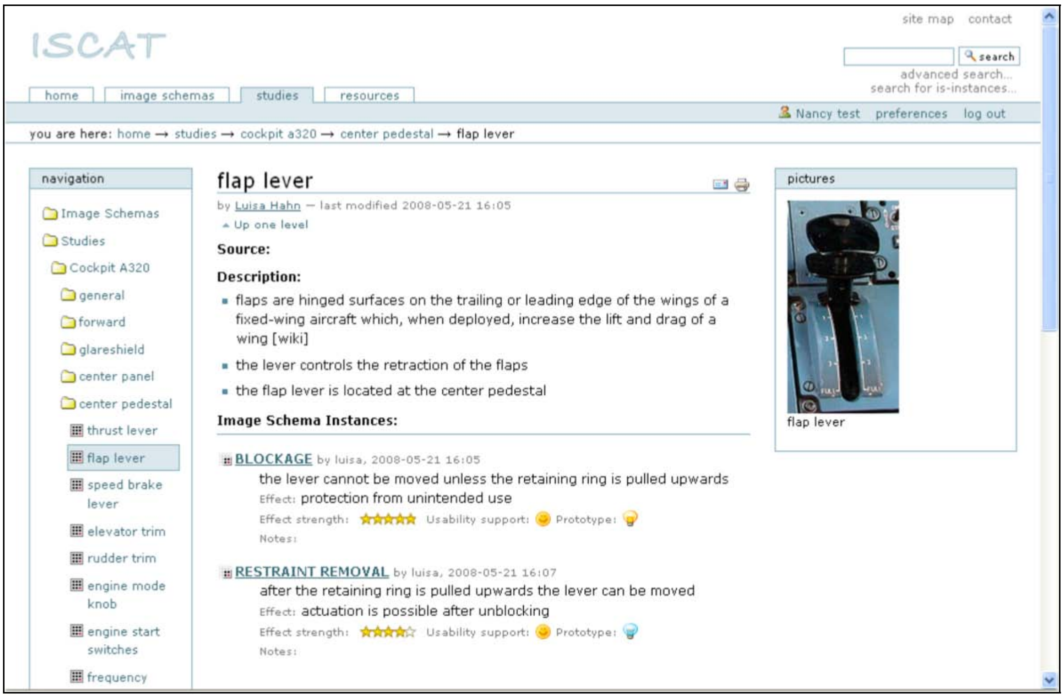


Figure 8.10. Example screen from the ISCAT database, page view of the user interface study of an Airbus 320

In addition, users have the ability to search the database via the search box provided in the upper right corner. Also, specific searches for image-schema instances are possible, the user can set more specific filters, e.g. on specific user interfaces, effect strength, and so on. Further, the advanced search offers possibilities to search for specific items like pictures, metaphorical extensions, etc. The database has an added user management system, so that new users can become members of the ISCAT community and contribute to the contents of the database.

8.2.2 Conclusion and Outlook

The image-schema database ISCAT seems a good starting point for providing the human-technology-research and design communities with an information repository on image schemas. Image-schema definitions can help to enhance the agreement between coders and reduce subjectivity when extracting image schemas during a context-of-use analysis. Examples of image-schema instantiations could inspire user interface designers in designing future solutions. Both, linguistic data and user interface examples, could contribute to new insights in user interface design.

Although such a database is practical, its users need to be reminded that the image schemas extracted from user interfaces (as well as the metaphorical extensions from linguistics) reflect the subjective view of the respective analyst. Thus, the examples in the database should be seen as a repository of hypotheses for researchers and heuristics for designers.

The accumulation of a large number of image-schema instances leads to the emergence of new research questions. The database could be used for mining general rules of image-schema application. Thus, implicitly applied design rules can be detected across different user-interface domains and can be subjected to further research. Early analyses of the current set of over 600 image-schema instances reveal

- rules of image-schema co-occurrences (e.g. BLOCKAGE needs to be followed by RESTRAINT REMOVAL, ATTRACTION is resulting in DIVERSION);
- image-schema transformation rules (e.g. UP-DOWN is readily substituted by FRONT-BACK relations);
- typical problems (e.g. UI elements that belong to the same task are often FAR away from each other without communicating their relation via a LINK or common CONTAINER image schema).

The database – tool and content – is open for further improvements. If designing with image schemas should be taken up by industry, a new version must meet the specific requirements of design practitioners. This version would feature better interconnections between the two parts of the database (linguistic and UI data). It would also provide better community support. The database could profit from including more kinds of data, for example data on typical user tasks and their associated image schemas. Other information could include typical constraints in using specific instantiations of image schemas that are determined by differing contexts of use. Finally, a formal evaluation of ISCAT will be necessary to determine its usability and general value for its users.

8.3 Practicability Revisited

The aim of this chapter was to assess the practicability of applying image schemas to human-centred design. A proof-of-concept study has been undertaken, issues of cost and benefits of using image schemas in various phases of the design process have been considered, and a database as a tool for supporting designers and researchers interested in applying image schemas was introduced. How can, from these and the studies reported in previous chapters, the 'practicability' of image schemas be judged?

Following a standard design process, the success of a method would be evaluated against its requirements. A set of requirements for methodological support in design is provided in a review article by Schmidt-Kretschmer and Blessing (2006). According to their analysis, methodology support in design should

- require as little effort for learning and training as possible,
- be easy to use,

- solve problems “in no time”,
- produce convincing results for complex problems,
- structure work sequence,
- be integrated in the existing design environments,
- support teamwork as well as individual work, and
- support compatibility of methods.

Although these requirements do not provide the specific thresholds that have to be met (e.g. what amount of time for problem-solving would be acceptable?), some general comments concerning these requirements can be made based on all seven studies that investigated various aspects of applying image schemas to user interface design.

Require as little effort for learning and training as possible. Although image schemas are a new concept in user interface design, they are easy to learn and easy to apply. Study 5, for instance, showed that 30 minutes were sufficient to explain not only the general concepts of image-schema theory but also a number of rather abstract FORCE image schemas to the novice participants of a workshop. This introduction enabled the participants to successfully relate their experiences of using products to the given image-schema categories. Study 5 also showed that, although spontaneous use of image schemas seems to be easily established, more training than 30 minutes may be required to enhance inter-rater agreements to a satisfactory level. A slightly longer training could be useful to capture the more subtle aspects of image-schema definitions and the interrelationships between different image schemas.

Be easy to use. Image schemas could be readily applied to usage scenarios as shown in study 5. Studies 6 and 7 showed that image schemas can be easily extracted from several aspects of the context-of-use and easily applied to generating design ideas. Image schemas are also easily extracted from user interfaces (studies 6 and 7). Problems with unclear terminology or unclear representations of image schemas are addressed by the ISCAT database. Practitioners can look up image-schema definitions and a collection of primary metaphors that until previously were scattered across the literature. Practitioners can also find a number of image-schema instances that help them in developing UI design solutions.

Solve problems “in no time”. Once image schemas are specified in the requirements, designers can actually save time when applying them, because they have a guidance for developing concrete design solutions from the requirements. Thus, image schemas help in bridging the design gap that is larger, when requirements come in the traditional way without image-schema guidance. Yet, quantitative studies are needed to explore exactly how much effort is saved by applying image schemas throughout a design process.

Produce convincing results for complex problems. Studies 1 to 4 have shown that the application of metaphorical extensions of image schemas can lead to large effects on the satisfaction of users, to small to medium effects on the mental efficiency, and that it can reduce error rates by as much as 100%. In study 7, image schemas were applied to the more complex context of redesigning a business application. The user evaluation showed that users rated the image-schematically designed UI prototypes better than their current system. Depending on the interaction paradigm (graphical or hybrid user interface), the evaluation varied concerning the practical and hedonic aspects of the user interfaces. Convincing results involving image schemas thus seem possible, but further studies need to extend the generalisability of the findings beyond the usage contexts investigated so far.

Structure work sequence. Image schemas are coupled to the general sequence of steps in a user-centred design process and can be applied to each. As a metalanguage they provide a holistic view of the design process.

Be integrated in the existing design environments. From the beginning on, the research done for this work had the objective to blend in image schemas with current design practices of a user-centred design lifecycle. Image schemas can be easily integrated in the methods already applied and cause little extra work. Further studies have to quantify the exact amounts of costs and benefits attached to applying image schemas within the common human-centred design framework.

Support teamwork as well as individual work. The image-schematic approach supports both types of work. Team-work should be preferred, when analysing image schemas from the context-of-use, as more raters will deliver more reliable (study 5) and more complete (study 6) results than a single rater can. Image schemas are also applicable to the division of labour found in practice, between a user-research team and a user-interface-design team. If image schemas are documented in the results of the context-of-use analysis and the requirements, the hand-over between the teams should be easily accomplished.

Support compatibility of methods. Image schemas are easily compatible with different methods of analysis and solution finding (e.g the morphological box as shown above). A common approach to producing design solutions, for instance, is the application of style guides (e.g. Microsoft Press, 1995) or of collections of design patterns (e.g. Tidwell, 2006; Yahoo, 2008) that in the analyses for the ISCAT database have shown to be easily compatible with and translatable into the image-schema metalanguage.

This discussion shows that image schemas have the potential to fulfil each of these requirements. Further research, however, needs to establish the concrete levels of achievement required of each of these aspects. For instance, what amount of training would user interface designers be willing to engage in to learn how to use image schemas? Follow-up studies could then specify whether these quantified goals can be reached with image-schema methodology.

9 Insights for User Interface Design

The central goal of this work is to determine whether image schemas can be a suitable guidance for designing intuitive use. To form a conclusion, it is now time to tie together the threads of the previous chapters and to summarise the discussion on intuitive use, image-schema theory, and the utility of image schemas in user interface design. Contrasting image schemas with the current guidance on intuitive use will show their value for designing intuitive use.

9.1 The Goal: Intuitive Use

Before the value of image schemas for designing intuitive use can be investigated, the goal needs to be clear. This means a definition of intuitive use must exist and suitable forms of measurement need to be derived. From the review of the literature and of previous investigations into the concept of intuitive use, the following definition of intuitive use was derived: Intuitive use is the extent to which a product can be used by subconsciously applying prior knowledge, resulting in an effective and satisfying interaction using a minimum of cognitive resources.

This definition includes preconditions for intuitive use as well as measurable consequences of intuitive use. Regarding the first precondition of subconscious information processing, the model of the human information processor (Rasmussen, 1986) gives some guidance on the nature of conscious and subconscious information processing and their interrelations. This model was enhanced in this work to accommodate different forms of conscious and subconscious knowledge representation. One form of subconscious knowledge representation are image schemas, which are also able to influence the development and processing of more conscious forms of representation. A process of contention scheduling was assumed that determines the activation and selection of specific image schemas.

To explain the second precondition for intuitive use, applying prior knowledge, a model of knowledge sources was devised. The model helps to distinguish different sources of knowledge referred to when talking about intuitive use. It was concluded that novices and heterogeneous user groups can be better reached by exploiting knowledge from the sensorimotor level of the model, because this type of knowledge is more basic than knowledge from the cultural or expertise levels. As image-schematic representations draw from sensorimotor experience, they were expected to be applicable to designing for large and heterogeneous user groups.

The consequences of intuitive use can be determined by measuring the effectiveness, satisfaction, and mental efficiency in use. Viewing intuitive use as a sub-concept of usability helps to derive useful operationalisations of the consequences of intuitive use. Like usability, intuitive use can be measured in terms of effectiveness and satisfaction. With regard to efficiency, it specialises on mental efficiency – in contrast, e.g. to motor effort or overall cost.

Although the given definition of intuitive use makes the concept manageable in theoretical terms, it is still difficult to demonstrate the subconscious processing of prior knowledge in practice. Researchers need to be cautious, because, while mental efficiency may be a necessary cue for subconscious processing, it is not a sufficient one. That is because differences in mental efficiency can also be detected between two conscious processes. The experiments on metaphorical extensions of image schemas in this work (studies 1 to 4) determined differences in mental efficiency that are assumed to occur due to the automatic triggering of metaphorical mappings. The data matches the expected consequences of intuitive use (higher mental efficiency for metaphor-congruent mappings), but proof is lacking that this is caused by the precondition of subconscious processing. But the results do not contradict subconscious processing, either.

The question is, whether proving the subconscious use of prior knowledge is (a) possible and (b) necessary. First, it is possible to obtain better indicators of subconscious processes using other subjective, objective, or physiological methods than applied here (e.g. Greenwald, McGhee & Schwartz, 1998; Seth, Dienes, Cleeremans, Overgaard, & Pessoa, 2008; Ziori & Dienes, 2006). Proving subconscious processing is more difficult than assessing differences in mental efficiency, though, and it often requires very sophisticated experimental techniques, rendering the studies artificial.

Second, is it necessary to prove the ‘subconscious application’ of knowledge? The answer is a theoretical ‘yes’ – and a practical ‘no’. In theoretical terms, this definition gives a clear distinction on what intuitive use is, i.e. every usage that happens below the level of consciousness. The preconditions require subconscious use of prior knowledge as this provides the theoretical gold standard for reducing mental load. Consequently, all guidance for designing intuitive use must be able to show how they can support subconscious processing. Formulating the preconditions in this way also guides the search for new design guidance. Image schemas, for example, would not have been found without the focus on subconscious processing. Other areas worth looking at can be marketing and advertising, as they rely heavily on the subconscious processing of information.

In practise, however, it is difficult to prove that prior knowledge is used exclusively subconsciously. During human-technology interaction, conscious and subconscious processes are intertwined. Even an entirely skill-based behaviour needs attentional checks that, if these detect something going wrong, hand over control to the conscious processor. In the choice reaction tasks of studies 1 to 4, the preparation of the action may have been subconscious, but the actual choice which button to press is still associated with consciousness. It is difficult to see how these processes can be disentangled. So, defining the consequences of intuitive use as likely outcomes and as practicable operationalisations relieves the burden of evaluation for designers. In practise, operationalising subconscious use in terms of mental efficiency, flanked by effectiveness and satisfaction as the other outcomes of intuitive use, will suffice to fulfil most demands. Further research, however, must strive to find better methods for detecting and measuring subconscious information processing and thus reconcile the theory and practise of designing for intuitive use.

9.2 Theoretical Foundation of Image Schemas

Image-schema theory originates in linguistic research and philosophical enquiry. Researchers in these disciplines claim that image schemas and their metaphorical extensions derive from sensorimotor experience, that they are a form of subconscious knowledge representation, and that they are metaphorically extended to abstract target domains. Despite such claims, no empirical evidence – beyond linguistic findings – is given in the original accounts. Recent psychological research provides evidence for the psychological reality of image schemas.

Concerning the claims on the sensorimotor origin of image schemas and their subconscious processing, developmental psychologists provide an account of the origins and the representation of image schemas in the infant's mind (Karmiloff-Smith, 1995; Mandler, 1992, 2004). A model of representational levels (procedural to explicit-verbal) was derived in this work that shows how image schemas as subconscious forms of processing interact with other representational formats within the dynamic world model of the human information processor of Rasmussen (1986).

The connection between automatic image-schema activation and sensorimotor processes has also been shown in the adult brain. Image schemas are automatically activated in language comprehension and interfere with or facilitate perception and motor action. Apart from language, factors contributing to the activation of image schemas are geometry, topological relation, function, frame of reference, and temporal aspects. Image schemas are also plausible in the light of neuroscientific evidence and theory (e.g. perceptual symbol systems, Barsalou, 1999).

Via metaphorical extensions, image schemas are involved in understanding abstract domains. The invariance hypothesis states that image schemas can provide the constraints for mapping larger domains of knowledge in metaphor. Experimental studies provide evidence that metaphors are conceptual instead of purely linguistic phenomena. Image schemas are triggered by non-verbal primes and influence perception and action in reaction to abstract concepts. Only few studies exist, however, that completely exclude non-linguistic stimuli.

From the review of the previous studies in linguistics and psychology, several conclusions can be drawn concerning the use of image schemas in user interface design. First, the evidence suggests that image schemas are a subconscious form of knowledge. As such they fulfil the defined preconditions of intuitive use, and applying them in user interfaces should enhance intuitive use.

Second, image schemas form primary metaphors. It was shown that primary metaphors are not restricted to the linguistic domain and can be triggered automatically. Applied to user interfaces, this means that image schemas are able to match the demand for solutions capable of mapping physical/spatial controls to abstract information.

Third, image schemas and their metaphorical extensions are ubiquitous. They can be found in different representational formats and media, such as language, drawings, and gestures. Image schemas are also found in languages across different cultures providing the underlying structure of spatial expressions and linguistic metaphors. An important consequence for user interface design is, that image schemas can be used for hardware and software user interfaces alike and designs informed by image-schema theory can be applied to designing for large and heterogeneous user groups. However, there are also variations of image-schema usage among cultures. Further research needs to determine the underlying sources for variation and stability, including non-linguistic data sources.

Fourth, the invariance hypothesis is a very promising idea, although concrete psychological evidence is rare. If the invariance hypothesis holds, image schemas can be used to describe users' tasks and mental models. Thus, image schemas can work as a metalanguage that helps in bridging the gap between user requirements and the design of user interfaces that are intuitive to use. Several of these promises have been addressed in the studies reported here.

9.3 Image Schemas in User Interface Design

The empirical studies in this work set off to test the application of image schemas and their metaphorical extensions in user interface design. The aim was to extend previous studies that were either not concerned with user interfaces or that applied image schemas only to selected phases of the design process. In particular, the focus here was on the verification of primary metaphors in user interfaces, and the reliability and practicability of image schemas as a metalanguage for user interface design.

9.3.1 Primary Metaphors in User Interfaces (Studies 1 to 4)

Image schemas were expected to enhance the intuitive use of user interfaces and their extensions to primary metaphors were expected to meet the demand for conveying abstract information at the user interface. These promises were tested in four experiments. In these, metaphor-congruent user interfaces were compared to metaphor-incongruent user interfaces and intuitive use was measured in terms of effectiveness (errors), mental efficiency (response times controlled for non-mental processes), and satisfaction (suitability judgments).

The results in general confirm that conceptual metaphors can enhance intuitive use. Metaphor-congruent user interfaces can lead to higher effectiveness, mental efficiency, and satisfaction than metaphor-incongruent user interfaces. However, the experiments also revealed that these effects can be modulated by characteristics of the task and of the user interface.

Studies 1 and 2 investigated metaphors of the image schemas UP-DOWN and LEFT-RIGHT, involving the target domains of valence (in study 1 differentiated into virtue and quality), quantity (both studies), and status (study 1 only). The overall results show that mental efficiency and satisfaction are higher for vertical button and slider arrangements that are metaphor-congruent than for metaphor-incongruent arrangements. The comparison with horizontal button and slider arrangements shows that these results are not due to the order in which the labels were read. A ceiling effect in effectiveness was observed; thus no meaningful comparisons could be made for this measure.

Varying the experimental conditions had differential effects on the processing of single metaphors. In study 1, the context of hotel evaluation emphasised valence ratings and deemphasised the target domains of quantity and status, so that for the latter no differences in mental efficiency were found. Removing the specifics of the context and introducing sliders as instances of the SCALE image schema in study 2 emphasised the target domain of quantity and deemphasised the domain of valence. This could be seen in the results for MORE IS UP mappings and in the reversal of the GOOD IS LEFT mapping to GOOD IS RIGHT in the horizontal condition. Further research needs to disentangle the effects of context and of instances of other image schemas present in the user interface.

Studies 3 and 4 investigated two contrasting metaphors of the NEAR-FAR image schema: CONSIDERED IS NEAR and SIMILAR IS NEAR. The results show that, in a similarity judgement task, the metaphor SIMILAR IS NEAR is more beneficial than the metaphor CONSIDERED IS NEAR. This result contradicts established usability guidelines like the proximity-compatibility principle. A limitation of this finding is that it could only be established under a moderately increased cognitive load on the user. The effect was found for mental efficiency and effectiveness, but not for satisfaction. Further research is required to discern to what extend this result is dependent on the type of task given to the participants.

The results of these studies show that the contribution of image schemas to the field of user interface design is threefold. First, studies 1 and 2 showed that primary metaphors can go beyond documented population stereotypes. More specifically, these studies added to the well-known mappings of MORE IS UP / MORE IS RIGHT the mappings GOOD IS UP / GOOD IS RIGHT. Second, these studies pointed out the differential utility of the same population stereotypes in varying contexts and varying instances of other image schemas present. Image schemas thus can help in describing and investigating differential effects that were not known before. Third, the results of study 4 showed that there are contexts in which established usability principles, such as the proximity compatibility principle, are not valid and the results can be better explained in terms of another primary metaphor.

Several hypotheses arise from the results of these studies that should be verified by further research:

- The effects of metaphorical mappings are moderated by other image-schema instances that are simultaneously present in the user interface (cf. the image schema SCALE in study 2).
- The task influences which metaphorical mapping is more beneficial in the case of conflicting predictions (cf. study 4).
- Physical mappings are stronger than metaphorical mappings (cf. the results on physical versus metaphorical UP-DOWN in studies 1 and 2).
- The differences between metaphor-congruent and incongruent mappings are expected to increase with mental workload (cf. the differences between study 3 and 4).

Finally, the broad range of image-schematic metaphors documented by cognitive linguistics provides a repository of potentially useful metaphorical mappings awaiting their test in a user interface context. For a balance between rigour and relevance, such studies should be both, lab-based and field-based.

9.3.2 Reliability of Image Schemas as a Design Language (Studies 5 and 6)

Another promise of image-schema theory derives from the invariance hypothesis. It holds that image schemas can be used as a common language to describe the structure of users' tasks and mental models and to use these descriptions to derive user interface designs. A precondition for using image schemas in a design language is that designers agree on which image schemas to associate with usage scenarios, user utterances, or user interface elements. Studies 5 and 6 were conducted to estimate the inter-rater reliabilities using an image-schema vocabulary.

Applying the FORCE image schemas in study 5 yielded moderate inter-rater reliabilities, $\kappa = .59$. Using the full range of image schemas across four components of use in study 6 yielded substantial inter-rater reliabilities, $\kappa = .66$ on average. Of the four components of the context of use, sufficiently high reliabilities were obtained for user interface screens ($\kappa = .94$) and users' utterances ($\kappa = .68$). Only moderate reliabilities were obtained for task steps ($\kappa = .40$) and steps of the user-system interaction ($\kappa = .45$). The overlap in the prevalence of image-schema categorisations between two raters was 75% on average – a good result compared to an average rate of 22% for usability problems found in expert usability evaluations.

When the assigned image schemas were compared to the intended image-schema assignments obtained in a pre-test, the agreement with the intended assignments was substantial for single raters and reached almost perfect agreement using two and more raters (study 5). To increase the validity of image-schema assignments, therefore, it is suggested to use at least pairs of raters.

In conclusion, the reliability of image-schema assignments depends on the type of information coded, the number of raters, and the image schemas used for evaluation. To enhance reliability scores, it is suggested to provide better access to context-of-use information for raters, to establish categorisation rules, and to train raters on image-schema definitions and categorisation rules. Further research needs to address the validity of rating rules and the effectiveness of rater trainings.

Although there is room for improvement in inter-rater reliabilities, these early results are promising. Using image schemas in the design process, therefore, can be encouraged – at least when designers are aware of the limitations and implications.

9.3.3 Practicability of Image Schemas as a Design Language (Study 7)

If image schemas are to be used as a design language, it must be feasible to apply them throughout the human-centred design cycle and the benefits and costs of application need to be favourable. A first exploration of these issues was conducted in study 7.

The study established that image schemas could be used for analysing the context of use, specifying requirements, producing design solutions, and evaluating user interfaces. Image schemas can be used to capture important aspects of the context of use such as the users' mental model of the task, the task structure, and the appearance and interaction with the current software system. The extracted image schemas can then be used to augment the requirement specification to provide the structure of the future user interface. When producing design solutions, image schemas facilitate the transfer from abstract requirements to concrete designs. Image-schematic design solutions can be devised individually for each requirement using the morphological box as a tool. The solutions then need to be selected and synthesised into an overall solution. In study 7, this approach was used to design a concept of a more traditional graphical user interface and a novel hybrid tangible-graphical user interface from the same set of requirements. Image schemas thus constrain the possible design space but leave the concrete instantiation in hardware or software to the creativity of the designer. User evaluations showed that the prototypes designed with image schemas lead to higher satisfaction scores regarding pragmatic quality, hedonic quality, and attractiveness than the current user interface solution.

The largest cost lies in the analysis phase, when image schemas must be extracted from the different components of the context of use. The largest benefit can be gained in the design phase, when the designer produces design solutions to meet the requirements. Here, image schemas prescribe the structure of the user interface and, at an abstract level, free the designer from applying pre-configured solutions.

To summarise, applying image schemas can provide a way to design more intuitive user interfaces. Further studies, however, need to go beyond low-fidelity prototyping to assess the effectiveness and mental efficiency of the use of products that are designed using image schemas. More research is necessary to more precisely estimate the costs and benefits of using image schemas in the design process and to disentangle the contributions made by image schemas from the contributions made by using a structured design process as such.

To support research and the practical application of image schemas, the database ISCAT was developed as the first central repository of image-schema definitions, their metaphorical extensions to abstract domains, and their application to user interface design. Here, designers can look up image-schema definitions and metaphorical extensions. Researchers are provided with a database that holds the potential for discovering implicit rules of image-schema syntax, semantics, and pragmatics in user interfaces.

To conclude, image schemas seem to be easily understood and easily applied by designers. They produce convincing qualitative and quantitative results. As a design language, image schemas provide a holistic view of the design process and are easily compatible with other methods in user-centred design.

Topics for further research on the practicability of image schemas are

- quantifying the costs and benefits of using image schemas in user interface design (compared to common UCD methodologies);
- setting up a methodology to make learning and using image schemas easy for designers without compromising reliability;
- developing the ISCAT database into an even more easy-to-use and profitable tool for design;
- collecting best and worst practices of using image schemas in user interface design and documenting these in ISCAT as a source of inspiration to designers.

While study 7 has shown how image schemas can be used in a human-centred design process, there are other ways of applying image-schema theory that are summarised in the following section.

9.3.4 Applying Image-Schema Theory

The basic principle of user interface design discussed in chapter 1 showed the importance of designing user interfaces driven by the users' tasks and mental models instead of driven by the underlying technology. Depending on the design phase, image-schema theory provides different alternatives as to how this can be achieved.

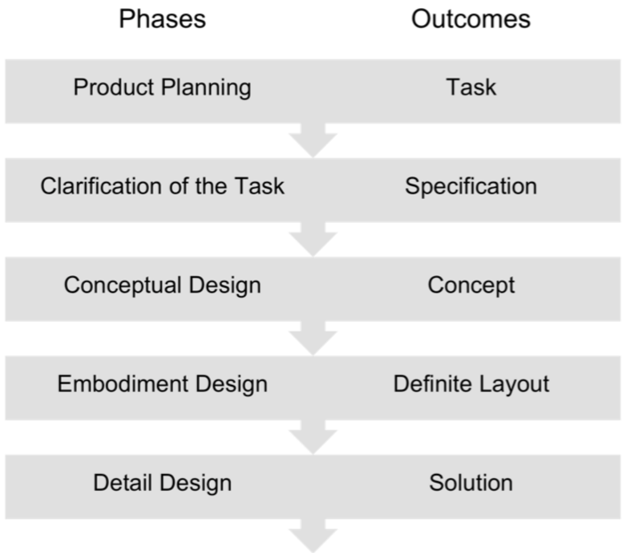


Figure 9.1. Phases and outcomes in the engineering design cycle (simplified, after Pahl, Beitz, Feldhusen, & Grothe, 2007)

The most valuable approach is to acknowledge that design for intuitive use is a prospective activity. Following a current engineering design methodology (*Figure 9.1*), design for intuitive use should already be considered in the phases of product planning and clarification of the task. An early focus on designing the user interface is reflected in the human-centred design cycle according to ISO 13407 (*Figure 4.1*) that can be easily integrated into the general product design cycle (cf. Abele et al., 2007; Mayhew, 1999). Study 7 shows how such a prospective design for intuitive use can be achieved by integrating image schemas into a standard human-centred design process according to ISO 13407.

But even if image schemas are not used as a descriptive language throughout a human-centred design cycle, image-schema theory can inspire designers. For example, in the phase of embodiment design (*Figure 9.1*), the definite layout of the product is established. The designer in this phase is supported by a number of guidelines, usually subsumed under the headline *Design for X*. In Design for Intuitive Use, the vast amount of primary metaphors documented in the literature can act as design guidelines. These guidelines can help in solving specific problems, e.g. in how to represent the concepts of *importance* or *valence* with physical/spatial user interface elements. Many of these primary metaphors are documented in the ISCAT database – a source designers can tap.

In case the metaphorical extensions that are needed are not yet documented, the theory can help to discern appropriate mappings. One strategy is to look at correlations in sensorimotor experience. For example, when mapping physical properties like speed or size to sound parameters like pitch, a designer may analyse correlations of these parameters in the real world. Considering cars passing on a road, for example, reveals correlations between speed and pitch;

considering large and small flying insects reveals correlations between size and pitch. According to the theory, these correlations in experience should have formed the conceptual metaphors HIGH PITCH IS FAST – LOW PITCH IS SLOW and HIGH PITCH IS SMALL – LOW PITCH IS BIG that can be exploited for devising the appropriate mappings in the user interface.

Another strategy is to look for such domain correlations in language. Here, the designer may find that spatial UP-DOWN is mapped to pitch as in *low bass* and *high-pitched shrieks* (HIGH PITCH IS UP – LOW PITCH IS DOWN). A third strategy is to watch how users behave in order to achieve a change at the user interface. Observing users changing the volume of a musical sound with gestures may reveal that they employ UP-DOWN or BIG-SMALL movements such that LOUD IS UP or LOUD IS BIG (cf. Bakker et al., 2009). Of course, when using these strategies, the usefulness of the derived mappings for achieving intuitive use needs to be verified – either informally in user validation sessions, or formally by experimental testing.

9.4 Image Schemas as a New Tool for Designing Intuitive Use

The goal of this work is to determine whether image schemas can be a suitable guidance for designing intuitive use. As there are already a number of current tools for designing intuitive use available, the value of adding image schemas to this toolbox needs to be determined.

9.4.1 Comparing Image Schemas to the Current Design Guidance

In chapter 2 an overview of the guidance for designing intuitive use was given. This guidance needs to be compared to image schemas in terms of theoretical foundations and meeting the demand for intuitive use.

Image Schemas and Gestalt Laws

Image schemas have been described as 'experiential gestalts' by Johnson (1987) in the sense that they are "coherent, meaningful, unified wholes within our experience and cognition. They are a principle means by which we achieve meaning structure" (p.41). Although gestalt laws and image schemas share the same theoretical basis, they differ in the levels of their mental processing. The effects described by gestalt laws are processed in the early stages of perception. Image schemas, in contrast, are involved in higher processing, building a bridge between perception and conception (Johnson, 1987; Robert, 1997). They are involved in understanding abstract domains, in language processing, etc. – to an extent that gestalt laws are not.

Image Schemas and Affordances

Image schemas have affordances, e.g. a CONTAINER has the affordance to open it. In terms of support for user interface design, the notion of affordance can only be a general approach to design rather than giving concrete suggestions. Also, affordances are restricted to communicate physical relationships. A graphical user interface element can have the perceived affordance to be draggable or clickable. Affordances cannot be used for conveying abstract information like *importance* or *happiness*. Image schemas, in contrast, give more concrete advice than affordances and the metaphorical extensions help in mapping physical UI elements to abstract data.

Image Schemas and Consistency

Image schemas and their metaphorical mappings fulfil the requirement for external consistency, because using them in user interfaces reflects basic experiences made outside the system. By applying image schemas and their metaphorical extensions to user interfaces, the designer exploits users' knowledge that has already been learned and automated. Users find familiar constraints and behaviour in the user interface and hence can use it with ease and high mental efficiency.

Nevertheless, to ensure internal consistency, image-schematic mappings should be consistently applied within a given user interface. Image schemas and consistency complement each other in that image schemas deliver concrete content to the notion of external consistency and can be the means to constrain variation for ensuring internal consistency.

Image Schemas and Compatibility

Compatibility refers to the mappings between displays and controls, between the user interface and the technical system, between the user interface and the mental model or the task of users. These mappings have often been described in physical terms, e.g. mappings of movements: UP-DOWN, FRONT-BACK, LEFT-RIGHT, ROTATION, or different ATTRIBUTES. Image schemas include and extend the common vocabulary of physical mappings – consider previously undiscussed mappings of CENTRE-PERIPHERY, IN-OUT, BLOCKAGE, or DIVERSION. Again, metaphorical extensions provide the possibility to not only encode direct physical-to-physical but also conceptual physical-to-abstract mappings.

Image Schemas and Population Stereotypes

Image schemas prescribe a number of mappings that are like population stereotypes – physical-to-physical mappings, e.g. UP IS FRONT – DOWN IS BACK, UP IS RIGHT – DOWN IS LEFT, as well as physical-to-abstract mappings involving metaphors. Image schemas expand the traditional notion of population stereotypes in three ways. First, image-schema theory provides a framework for grounding stereotypical mappings in everyday experience, while, until now, population stereotypes have been recorded without much theorising about their origins. Having a theory behind the notion of population stereotypes allows predicting and testing new hypothetical population stereotypes.

Second, the number of metaphorical extensions documented by cognitive linguists exceeds the number of documented population stereotypes by an estimated factor of ten. Here, many hypotheses and heuristics await their investigation as population stereotypes. Third, studies 1 to 4 provided performance data to estimate the impact of population stereotypes on the consequences of intuitive use, going beyond the traditional method of paper-and-pencil tests that at most deliver subjective preferences and thus satisfaction measures.

A tradition in the population stereotype literature is to look at the conflicting predictions made by different population stereotypes. Their resolution then can lead to new guidelines (e.g. Warrick's principle, cf. Proctor & Vu, 2006). Similar research needs to be done on metaphorical extensions as well. First steps are made in studies 3 and 4 above and they provide hypotheses on the task dependency of image-schematic effects.

Image Schemas and User Interface Metaphor

Many issues about using traditional user interface metaphors revolve around the correct mappings from a real-world source domain to the user interface. Most authors theorising about UIM are aware of conceptual metaphor theory – they use it as justification for why UIM should be intuitive to use (e.g. Barr, Biddle, & Noble, 2002; Madsen, 1994). But relying too much on real-world metaphors can also get designers stuck. Applying the invariance principle to reasoning about UIM can help. It claims that the mappings in rich metaphors are constrained by image schemas. Image schemas help in maintaining the essential structure of complex metaphors. At the same time they can free user interfaces from mimicking existing technological and cultural artefacts.

This can also solve the problem in which intuitive use is seen to hamper progress (Raskin, 1994; cf. chapter 2.1). Designing with image schemas and their metaphorical extensions builds on very basic representations in interacting with the world that hardly can become outdated. Image schemas are abstract and flexible enough to be adapted to every kind of UI design – hardware as well as software user interfaces. New technology is no excuse not to design for intuitive use.

9.4.2 Summary

In conclusion, image-schema theory can be considered a worthy member of the toolbox for designing intuitive use. The particular features that set image-schema theory apart from the other tools are:

- the invariance hypothesis: the idea that abstract domains are structured by image schemas;

- the concrete prescription of numerous physical-to-abstract mappings via direct metaphorical extensions of image schemas;
- the grounding of image schemas and their metaphorical extensions in basic sensorimotor experience making them meaning-bearing subconscious representations; and
- the mixture of theoretical approach and concrete predictions for user interface design.

9.5 Conclusion and Outlook

Should image schemas be recommend as a future tool for designing intuitive use? Given the theory and the results of the studies presented in this work, the answer is 'yes'. The studies have shown that the scope of image schemas, although developed in philosophy and cognitive linguistics, goes beyond phenomenological introspection and linguistic explanation. Image schemas are a cognitive phenomenon, a form of abstract sensorimotor knowledge representation in long-term memory. This plus their subconscious processing makes image schemas fit the defined preconditions of intuitive use. Although derived from sensorimotor knowledge, they also form the basic structure of abstract thought via primary metaphors or by providing the structure that is mapped between more complex domains (as the invariance hypothesis holds). The metaphorical extensions of image schemas are numerous. Image schemas are part of users' mental models and therefore are relevant for designing user interfaces that match these mental models. Using image-schematic metaphors opens up new ways to meet the demand for conveying abstract data at the user interface. Using image schemas as the vocabulary of a design language turns them into a useful tool for designing intuitive interaction.

The studies presented in this work show that applying image schemas and their metaphorical extensions to the design of user interfaces enhances measurements of intuitive use. Implementing user interfaces that are congruent with image-schematic metaphors leads to higher effectiveness, mental efficiency, and satisfaction in use than using user interfaces that contradict image-schematic metaphors. As a relatively fixed vocabulary of image schemas is established, it can be used as an abstract language for describing requirements and prescribing user interface design solutions. The studies found that image schemas can form a reliable description language for user interfaces and users' utterances, and they can be used as a practicable language in all stages of a human-centred design cycle.

However, the studies have also shown that we are only at the beginning of understanding the use of image schemas in user interfaces. Questions have arisen that need to be addressed by further research and that were listed elsewhere in this chapter. These questions can be summarised to form larger areas that have to be addressed in studying image schemas and their metaphorical extensions – within and outside a user interface design context.

First, more research needs to be undertaken regarding the fundamental claims of image-schema theory. The conceptual nature of a rich set of linguistic findings needs to be verified by further experimental research using non-linguistic stimuli. The existing list of image schemas needs to be investigated regarding completeness, conceptual reality, and interdependencies of image schemas.

Second, the concept of intuitive use puts a focus on the subconscious use of prior knowledge, which has not been a focus in previous research in human-technology interaction. Because presently it is difficult to confirm subconscious processing empirically, it is necessary to further develop the theory and the methodology of measuring intuitive use in a practicable and valid way.

Third, to explore image schemas as a metalanguage in design, further studies should apply them in design projects and determine costs and benefits of image schema usage more precisely than could be done here. It will also be necessary to develop full prototypes on which all components of intuitive use can be assessed, including user performance data. To help designers apply image schemas as a metalanguage, further research needs to go into enhancing the reliability of applying the language and into further developing guiding tools such as the ISCAT database.

Fourth, research should look at the features of user interface elements that activate image schemas in the users' minds. It is necessary to determine how these activations depend on the task or other image-schema instantiations present in the user interface. Further, the validity of the consciously assigned image-schema instantiations in a context-of-use analysis needs to be verified, i.e. whether these image schemas are actually relevant to the users in solving their tasks.

Fifth, a research topic left out in the present work is the universality of image schemas. As image schemas and primary metaphors arise from basic sensorimotor experience that is common to different people and cultures, image schemas and primary metaphors should be ubiquitous and universally valid. Although linguistic findings support this view, studies going beyond linguistic data are rare, and studies on the application in user interfaces have not been considered yet. However, as one promise of image schemas is to meet the demand for supporting large and heterogeneous user groups, addressing this research area is essential for the application of image schemas in design for intuitive use.

To conclude, although image-schema theory is a young field of investigation, its promises for designing intuitive use are stimulating. This work has explored some of the relevant questions and pointed out many more. The results show that image schemas are worth pursuing as design guidance for intuitive use. Their great promise and vast scope of application in user interface design should encourage researchers and practitioners to carry on.

References

Antle, A. N., Corness, G., & Droumeva, M. (2009). What the body knows: Exploring the benefits of embodied metaphors in hybrid physical digital environments. *Interacting with Computers, 21*(1-2), 66-75.

Antle, A. N., Droumeva, M., & Corness, G. (2008). Playing with the sound maker: do embodied metaphors help children learn? In *Proceedings of the 7th international conference on Interaction design and children, IDC '08* (pp. 178–185). New York: ACM.

Abele, P., Hurtienne, J., & Prümper, J. (2007). *Usability Management bei SAP-Projekten. Grundlagen, Vorgehen, Methoden*. Wiesbaden: Vieweg.

Agor, W. H. (1986). *The logic of intuitive decision making: A research-based approach for top management*. New York: Quorum Books.

Appelle, S. (1972). Perception and discrimination as a function of stimulus orientation: The "oblique effect" in man and animals. *Psychological Bulletin, 78*(4), 266-278.

Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge: Cambridge University Press.

Baillargeon, R., Kotovsky, L., & Needham, A. (1995). The acquisition of physical knowledge in infancy. In D. Sperber, D. Premack & A. J. Premack (Eds.), *Causal Cognition: A Multidisciplinary Debate* (pp. 79-116). New York: Oxford University Press.

Baillargeon, R., & Wang, S. (2002). Event categorization in infancy. *Trends in Cognitive Sciences, 6*(2), 85-93.

Baldauf, C. (1997). *Metapher und Kognition. Grundlagen einer neuen Theorie der Alltagsmetapher.* Frankfurt am Main: Peter Lang.

Bakker, S., Antle, A. N., & van der Hoven, E. (2009). Identifying embodied metaphors in children's sound-action mappings. In P. Paolini & F. Garzotto (Eds.), *Proceedings of the 8th International Conference on Interaction Design and Children* (pp. 140-149). Como, Italy: ACM.

Bargh, J.A., & Chartrand, T.L. (1999). The Unbearable Automaticity of Being. *American Psychologist, 54*(7), 462-479.

Bargh, J. A., & Ferguson, M. J. (2000). Beyond behaviorism: On the automaticity of higher mental processes. *Psychological Bulletin, 126*(6), 925-945.

Barr, P., Biddle, R., & Noble, J. (2002). A taxonomy of user interface metaphors. Technical Report CS-TR-02-11. Wellington, New Zealand: Victoria University of Wellington.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*(4), 577-609; 637-660.

Bastick, T. (2003). *Intuition: Evaluating the construct and its impact on creative thinking*. Kingston, Jamaica: Stoneman and Lang.

Bergen, B., Lindsay, S., Matlock, T., & Narayanan, S. (2007). Spatial and linguistic aspects of visual imagery in sentence comprehension. *Cognitive Science, 31*, 733-764.

Berner, W. (2006). *Angst: Die wichtigste Emotion in Veränderungsprozessen*. Retrieved May 16, 2008, from http://www.umsetzungsberatung.de/psychologie/angst.php

Bias, R. G. (1994). The Pluralistic Usability Walkthrough: Coordinated Empathies. In J. Nielsen & R. L. Mack (Eds.), *Usability Inspection Methods* (pp. 63-76). New York: Wiley.

Blackler, A. (2006). *Intuitive interaction with complex artefacts*. Doctoral Thesis. Brisbane: Queensland University of Technology.

Blackler, A., & Hurtienne, J. (2007). Towards a unified view of intuitive interaction: definitions, models and tools across the world. *MMI-Interaktiv, 13*, 36-54.

Blackler, A., Popovic, V., & Mahar, D. (2007). Empirical investigations into intuitive interaction: a summary. *MMI-Interaktiv, 13*, 4-24.

BNC (2007). *The British National Corpus, version 3*. Retrieved July 12, 2009, from http://www.natcorp.ox.ac.uk/

Boers, F. (1999). When a bodily source domain becomes prominent: The joy of counting metaphors in the socio-economic domain. In R. W. Gibbs & G. J. Steen (Eds.), *Metaphor in Cognitive Linguistics. Selected papers from the Fifth International Cognitive Linguistics Conference Amsterdam, July 1997* (pp. 47-56). Amsterdam, Philadelphia: John Benjamins.

Boroditsky, L. (2000). Metaphoric structuring: Understanding time through spatial metaphors. *Cognition, 75*(1), 1-28.

Boroditsky, L. (2001). Does Language Shape Thought? Mandarin and English Speakers' Conceptions of Time. *Cognitive Psychology, 43*(1), 1-22.

Boroditsky, L., & Ramscar, M. (2002). The roles of mind and body in abstract thought. *Psychological Science, 13*(2), 185-188.

Bortz, J., & Döring, N. (2002). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Berlin: Springer.

Bortz, J., & Lienert, G. A. (1998). *Kurzgefaßte Statistik für die klinische Forschung. Ein praktischer Leitfaden für die Analyse kleiner Stichproben*. Berlin: Springer.

Bowden, E. M. (1997). The Effect of Reportable and Unreportable Hints on Anagram Solution and the Aha! Experience. *Consciousness and Cognition, 6*(4), 545-573.

Bowdle, B., & Gentner, D. (2005). The career of metaphor. *Psychological Review, 112*(1), 193–216.

Bowers, K. S., Regehr, G., Balthazard, C., & Parker, K. (1990). Intuition in the Context of Discovery. *Cognitive Psychology, 22*(1), 72-110.

Brewer, W. F. (1987). Schemas versus mental models in human memory. In P. Morris (Ed.), *Modelling cognition* (pp. 187-197). Chichester: John Wiley and Sons.

Briell Marketing. (2008). *Sony Digital Printer, Snap Lab UPCR10L*. Retrieved May 31, 2008, from http://www.briell.com.au/sony/snap%20lab.htm

Casasanto, D. (2009a). Embodiment of abstract concepts: Good and bad in right- and left-handers. *Journal of Experimental Psychology: General*, *138*(3), 351-367.

Casasanto, D. (2009b). When is a Linguistic Metaphor a Conceptual Metaphor? In V. Evans & S. Pourcel (Eds.), *New Directions in Cognitive Linguistics* (pp. 127-145). Amsterdam: John Benjamins.

Casasanto, D., & Boroditsky, L. (2008). Time in the mind: Using space to think about time. *Cognition, 106*(2), 579-593.

Casasanto, D., Boroditsky, L., Phillips, W., Greene, J., Goswami, S., Bocanegra-Thiel, S., et al. (2004). How deep are effects of language on thought? Time estimation in speakers of English, Indonesian, Greek, and Spanish. In K. Forbus, D. Gentner & T. Regier (Eds.), *Proceedings of the 26th Annual Meeting of the Cognitive Science Society* (pp. 186–191). Mawah, NJ: Lawrence Erlbaum.

Casasanto, D., & Lozano, S. (2006). Metaphor in the Mind and Hands. In R. Sun & N. Miyake (Eds.), *Proceedings of 28th Annual Conference of the Cognitive Science Society* (pp. 142-147). Hillsdale, NJ: Lawrence Erlbaum Associates.

Casasanto, D., & Lozano, S. (2007). Meaning and Motor Action. In D. McNamara & G. Trafton (Eds.), *Proceedings of the 29th Annual Meeting of the Cognitive Science Society* (pp. 149-154). Austin, TX: Cognitive Science Society.

Casasanto, D., Phillips, W., & Boroditsky, L. (2003). Do we think about music in terms of space? Metaphoric representation of musical pitch. In R. Alterman & D. Kirsh (Eds.), *Proceedings of 25th Annual Conference of the Cognitive Science Society* (p.1323). Hillsdale, NJ: Lawrence ErlbaumAssociates.

CDC (2008). Centers for Disease Control and Prevention: International Influenza Surveillance. Retrieved January 14, 2008, from http://www.cdc.gov/flu/weekly/intsurv.htm

Chen, J. Y. (2007). Do Chinese and English speakers think about time differently? Failure of replicating Boroditsky (2001). *Cognition, 104*(2), 427-436.

Choi, S., & Bowerman, M. (1991). Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition, 41*(1-3), 83-121.

Cienki, A. (1993). The strengths and weaknesses of the left/right polarity in Russian: Diachronic and synchronic semantic analyses. In L. d. Stadler & C. Eyrich (Eds.), *Issues in Cognitive Linguistics* (pp. 299-325). Berlin, New York: Mouton de Gruyter.

Cienki, A. (1998). STRAIGHT: An image schema and its metaphorical extensions. *Cognitive Linguistics, 9*(2), 107–149.

Cienki, A., & Müller, C. (2008). Metaphor, gesture, and thought. In R. W. Gibbs (Ed.), *The Cambridge handbook of metaphor and thought* (pp. 483–500). Cambridge, UK: Cambridge University Press.

Clausner, T. C. (2005). Image schema paradoxes: Implications for cognitive semantics. In B. Hampe & J. E. Grady (Eds.), *From perception to meaning: image schemas in cognitive linguistics* (pp. 93-110). Berlin; New York: Mouton de Gruyter.

Clausner, T. C., & Croft, W. (1999). Domains and image schemas. *Cognitive Linguistics, 10*(1), 1-31.

Cobb, J.G. (2002, May 12). Menus Behaving Badly. *New York Times*. Retrieved from http://www.nytimes.com/2002/05/12/automobiles/menus-behaving-badly.html

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement, 20*(1), 37-46.

ControlGlobal. (2008). *New Products: Keypad improves intuitive use for drive controls.* Retrieved May 31, 2008, from http://www.controlglobal.com/vendors/products/2004/159.html

Cooper, A., & Reimann, R. (2003). *About face 2.0: The essentials of interaction design*. Indianapolis, IN: Wiley.

Cooper, R., & Shallice, T. (2000). Contention scheduling and the control of routine activities. *Cognitive Neuropsychology, 17*(4), 297-338.

Correa-Benningfield, M., Kristiansen, G., Navarro-Ferrando, I., & Vandeloise, C. (2005). Image schemas vs. "Complex Primitives" in cross-cultural spatial cognition. In B. Hampe & J. E. Grady (Eds.), *From perception to meaning: image schemas in cognitive linguistics* (pp. 343-366). Berlin; New York: Mouton de Gruyter.

Coventry, K. R., Carmichael, R., & Garrod, S. C. (1994). Spatial Prepositions, Object-Specific Function, and Task Requirements. *Journal of Semantics, 11*(4), 289-309.

Coventry, K. R., Prat-Sala, M., & Richards, L. (2001). The interplay between geometry and function in the comprehension of over, under, above, and below. *Journal of Memory and Language, 44*(3), 376-398.

Cowey, A., Small, M., & Ellis, S. (1994). Left visuo-spatial neglect can be worse in far than in near space. *Neuropsychologia, 32*(9), 1059-1066.

Crawford, L. E., Margolies, S. M., Drake, J. T., & Murphy, M. E. (2006). Affect biases memory of location: Evidence for the spatial representation of affect. *Cognition and Emotion, 20,* 1153–1169.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York: Putnam.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General, 122*(3), 371–396.

Deutsche Telekom. (2007). *Advancing Innovation.* Retrieved Feb 8, 2008, from http://www.laboratories.telekom.com/ipws/English/LabsGeneral/Strategie/Pages/default.aspx

Dewell, R. B. (2005). Dynamic patterns of CONTAINMENT. In B. Hampe & J. E. Grady (Eds.), *From perception to meaning: image schemas in cognitive linguistics* (pp. 369-393). Berlin; New York: Mouton de Gruyter.

Dixon, R. M. W. (1982). *Where have all the adjectives gone? and other essays in semantics and syntax.* Berlin, Amsterdam, NewYork: Mouton.

Djajadiningrat, T., Wensveen, S., Frens, J., & Overbeecke, K. (2004). Tangible products: redressing the balance between appearance and action. *Personal and Ubiquitous Computing, 8*, 294-309.

Dong, S. (2007). *Catalog of user interface elements for intuitive user interfaces.* Unpublished student project, Technische Universität, Berlin.

Dreyfus, H. L., Dreyfus, S. E., & Athanasiou, T. (1986). *Mind over machine: The power of human intuition and expertise in the era of the computer*. Oxford, UK: B. Blackwell.

Dubberly, H., Evenson, S., & Robinson, R. (2008). The Analysis-Synthesis Bridge Model. *interactions, 15*(2), 57-61.

Eugenio, B. D., & Glass, M. (2004). The Kappa Statistic: A Second Look. *Computational Linguistics, 30*(1), 95-101.

Eysenck, H. J. (1995). *Genius: The natural history of creativity*. Cambridge; New York: Cambridge University Press.

Eysenck, M. W., & Keane, M. T. (2005). *Cognitive psychology: A student's handbook* (5th ed.). Hove, UK; New York: Psychology Press.

Feist, M. (2000). *On In and On: An investigation into the linguistic encoding of spatial scenes.* Unpublished Dissertation, Northwestern University, Evanston, IL.

Fischbein, E. (1987). *Intuition in science and mathematics: An educational approach*. Dordrecht, Holland: Reidel.

Fischer, A. R. H. (1999). *Intuitive Interfaces: A literature review of the Natural Mapping principle and Stimulus Response compatibility*. Retrieved Oct 24, 2007, from http://www.idemployee.id.tue.nl/g.w.m.rauterberg/publications/NATMAPreport.pdf

Fischer, M. H., Warlop, N., Hill, R. L., & Fias, W. (2004). Oculomotor Bias Induced by Number Perception. *Experimental Psychology, 51*(2), 91-97.

Fitts, P. M. (1951). *Human Engineering for an Effective Air-navigation and Traffic-control System*. Columbus, OH: National Research Council, Division of Anthropology and Psychology, Committee on Aviation Psychology.

Fitzpatrick, J. A. (2001). *Go! buttons and arrow icons: How metaphors motivate conventions in web design*. Retrieved May 29, 2008, from http://www.obscure.org/~jafitz/lin106/web_paper/web_conventions.pdf

Frese, M., & Brodbeck, F. C. (1989). *Computer in Büro und Verwaltung: Psychologisches Wissen für die Praxis*. Berlin, Heidelberg: Springer-Verlag.

Garrod, S., Ferrier, G., & Campbell, S. (1999). In and on: Investigating the functional geometry of spatial prepositions. *Cognition, 72*(2), 167-189.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7*(2), 155-170.

Gentner, D. (2001). Spatial metaphors in temporal reasoning. In M. Gattis (Ed.), *Spatial schemas and abstract thought* (pp. 203–222). Cambridge, MA: MIT Press.

Gentner, D., & Markman, A. B. (1997). Structure Mapping in Analogy and Similarity. *American Psychologist*, *52*(1), 45-56.

Gevers, W., Reynvoet, B., & Fias, W. (2003). The mental representation of ordinal sequences is spatially organized. *Cognition, 87*(3), B87-B95.

Gibbs, R. W. (2005). The psychological status of image schemas. In B. Hampe & J. E. Grady (Eds.), *From perception to meaning: image schemas in cognitive linguistics* (pp. 113-135). Berlin; New York: Mouton de Gruyter.

Gibbs, R. W., Beitel, D. A., Harrington, M., & Sanders, P. E. (1994). Taking a stand on the meanings of stand: Bodily experience as motivation for polysemy. *Journal of Semantics, 11*(4), 231-251.

Gibbs, R. W., & Colston, H. L. (1995). The cognitive psychological reality of image schemas and their transformations. *Cognitive Linguistics, 6*(4), 347-378.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Giessner, S. R., & Schubert, T. W. (2007). High in the hierarchy: How vertical location and judgments of leaders' power are interrelated. *Organizational Behavior and Human Decision Processes, 104*(1), 30-44.

Gigerenzer, G. (2007). *Gut feelings: The intelligence of the unconscious*. New York: Viking.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9*(3), 558-565.

Glover, S., Rosenbaum, D. A., Graham, J., & Dixon, P. (2004). Grasping the meaning of words. *Experimental Brain Research, 154*(1), 103-108.

Glucksberg, S., & McGlone, M. S. (2001). *Understanding figurative language: From metaphors to idioms*. Oxford; New York: Oxford University Press.

Goldstone, R. L. (1994). An efficient method for obtaining similarity data. *Behavior Research Methods, Instruments, & Computers, 26*, 381-386.

Goschler, J. (2008). *Metaphern für das Gehirn. Eine kognitiv-linguistische Untersuchung.* Berlin: Frank & Timme.

Gould, J. D., & Lewis, C. (1985). Designing for usability – Key principles and what designers think. *Communications of the ACM, 28*(3), 300-311.

Grady, J. (1997a). *Foundations of meaning: Primary metaphors and primary scenes*. PhD dissertation, University of California at Berkeley.

Grady, J. (1997b). THEORIES ARE BUILDINGS revisited. *Cognitive Linguistics, 8*(4), 267-290.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*(6), 1464-1480.

Hadank, J. M., Allen, W. E., Bradbury, W. J., & Anderson, P. D. (1991*). Intuitive joystick control for a work implement*. US Patent 5002454. Retrieved Feb 08, 2008, from http://www.patentstorm.us/patents/5002454-description.html

Hahn, L. (2007). *Analyse eines Cockpits zur Erstellung eines Kataloges von Interaktionselementen.* Unpublished student project, Technische Universität, Berlin.

Halligan, P. W., & Marshall, J. C. (1991). Left neglect for near but not far space in man. *Nature, 350*, 498-500.

Hampe, B. (2005). Image Schemas in Cognitive Linguistics: Introduction. In B. Hampe & J. E. Grady (Eds.), *From perception to meaning: image schemas in cognitive linguistics* (pp. 1-12). Berlin; New York: Mouton de Gruyter.

Hampe, B., & Grady, J. E. (Eds.). (2005). *From perception to meaning: Image schemas in cognitive linguistics*. Berlin; New York: Mouton de Gruyter.

Harper, D., Fallon, S., Gaskell, K., Grundvig, J., Heller, C., Huhti, T., et al. (2005). *China* (Lonely Planet Country Guide). Hawthorn, Vic., Australia; Oakland, CA: Lonely Planet.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 139–183). Amsterdam: Nord Holland Press.

Haser, V. (2003). Metaphor in semantic change. In A. Barcelona (Ed.), *Metaphor and Metonymy at the Crossroads: A Cognitive Perspective* (pp. 171-194). Berlin, New York: Mouton de Gruyter.

Hassenzahl, M., Burmester, M., & Koller, F. (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In J. Ziegler & G. Szwillus (Eds.), *Mensch & Computer 2003. Interaktion in Bewegung.* (pp. 187-196). Stuttgart, Leipzig: B.G. Teubner.

Hayward, W. G., & Tarr, M. J. (1995). Spatial language and spatial representation. *Cognition, 55*(1), 39-84.

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. New York: Wiley.

Hertzum, M., & Jacobsen, N. E. (2003). The Evaluator Effect: A Chilling Fact About Usability Evaluation Methods. *International Journal of Human-Computer Interaction, 15*(1), 183-204.

Hespos, S. J., & Baillargeon, R. (2001). Reasoning about containment events in very young infants. *Cognition, 78*(3), 207-245.

Hinske, S. (2007). Determining the Position and Orientation of Multi-Tagged Objects Using RFID Technology. In *IEEE Pervasive Computing and Communications Workshops (PerCom '07)* (pp. 377-381). White Plains, NY: IEEE.

Holtzblatt, K., Wendell, J. B., & Wood, S. (2005). *Rapid contextual design: A how-to guide to key techniques for user-centered design*. San Francisco: Elsevier/Morgan Kaufmann.

Howell, D. C. (1997). *Statistical methods for psychology* (4th ed.). Belmont, CA: Duxbury Press.

Hung, Y. H., Hung, D. L., Tzeng, O. J.-L., & Wu, D. H. (2008). Flexible spatial mapping of different notations of numbers in Chinese readers. *Cognition, 106*, 1441–50.

Hurricane Software Inc. (2002). Legal Practice search tool. Retrieved May 31, 2008, from http://www.hurricanesoft.com/hslegal.jsp

Hurtienne, J., & Blessing, L. (2007). Metaphors as Tools for Intuitive Interaction with Technology. *metaphorik.de*(12), 21-52.

Hurtienne, J., Dinsel, C., & Sturm, C. (2009). *QUESI – Design and evaluation of a questionnaire for the subjective consequences of intuitive use.* Unpublished manuscript, Fachgebiet Mensch-Maschine Systeme, Technische Universität Berlin.

Hurtienne, J., & Israel, J. H. (2007). Image schemas and their metaphorical extensions: intuitive patterns for tangible interaction. In B. Ullmer, A. Schmidt, E. Hornecker, C. Hummels, R. J. K. Jacob & E. van den Hoven (Eds.), *TEI'07 First international conference on Tangible and embedded interaction* (pp. 127-134). New York: ACM.

Hurtienne, J., Israel, J. H., & Weber, K. (2008). Cooking up real world business applications combining physicality, digitality, and image schemas. In A. Schmidt, H. Gellersen, E. v. d. Hoven, A. Mazalek, P. Holleis & N. Villar (Eds.), *TEI'08. Second International Conference on Tangible and Embedded Interaction* (pp. 239-246). New York: ACM.

Hurtienne, J., Langdon, P., & Clarkson, P. J. (2009). Towards an Account of Sensorimotor Knowledge in Inclusive Product Design. In C. Stephanidis (Ed.), *Universal Access in Human-Computer Interaction: Addressing Diversity, Lecture Notes in Computer Science* (Vol. 5614, pp. 251–260). Berlin: Springer.

Hurtienne, J., Prümper, J. & Rötting, M. (2009). When enterprise resource planning needs software ergonomics: Some typical scenarios. In *Proceedings of the 17th World Congress on Ergonomics (IEA 2009)* [CD-ROM]. Beijing, China: IEA.

Hurtienne, J., Stößel, C., Sturm, C., Maus, A., Rötting, M., Langdon, P., & Clarkson, P. J. (2010). Physical gestures for abstract concepts. Inclusive design with primary metaphors. *Interacting with Computers, 22*(6), 475-484.

Hußlein, S., Hurtienne, J., Israel, J. H., Mohs, C., Kindsmüller, M. C., Meyer, H. A., et al. (2007). Intuitive Nutzung - nur ein Schlagwort? *design report, 07*(11), 26-27.

Ishii, H. (2008). Tangible bits: Beyond pixels. In A. Schmidt, H. Gellersen, E. v. d. Hoven, A. Mazalek, P. Holleis & N. Villar (Eds.), *TEI'08. Second International Conference on Tangible and Embedded Interaction* (pp. xv-xxv). New York: ACM.

Ishii, H., & Ullmer, B. (1997). Tangible bits: Towards seamless interfaces between people, bits and atoms. In S. Pemberton (Ed.), *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 234-241). New York: ACM.

ISO (1998). *ISO 9241-11: Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability*. Geneva: International Organization for Standardization.

ISO (1999). *ISO 13407: Human-centred design processes for interactive systems*. Geneva: International Organization for Standardization.

ISO (2006). *ISO 9241-110: Ergonomics of human-system interaction - Part 110: Dialogue principles*. Geneva: International Organization for Standardization.

IST (2006). *Coverage and main targets.* Retrieved Feb 8, 2008, from http://cordis.europa.eu/ist/about/coverage.htm

Jacobs, B. (2005). *Einige Berechnungsmöglichkeiten von Effektstärken.* Retrieved April 1, 2008, from http://www.phil.uni-sb.de/~jakobs/seminar/vpl/bedeutung/effektstaerketool.htm

Jäkel, O. (2003). *Wie Metaphern Wissen schaffen*. Hamburg: Dr. Kovac.

January, D., & Kako, E. (2007). Re-evaluating evidence for linguistic relativity: Reply to Boroditsky (2001). *Cognition, 104*(2), 417-426.

Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason*. Chicago: University of Chicago Press.

Johnston, J. R. (1988). Children's Verbal Representations of Spatial Location. In J. Stiles-Davies, M. Kritchevsky & U. Bellugi (Eds.), *Spatial Cognition* (pp. 195-205). Hillsdale, NJ: Lawrence Erlbaum.

Karem, A. E. (2001). *Intuitive vehicle and machine control.* International Patent WO/2001/088648. Retrieved Feb 8, 2008, from http://www.wipo.int/pctdb/en/wo.jsp?WO=2001%2F088648&IA=WO2001%2F088648&DISPLAY=STATUS

Karmiloff-Smith, A. (1995). *Beyond modularity: A developmental perspective on cognitive science*. Cambridge, Mass.: MIT Press.

Kaschak, M. P., Madden, C. J., Therriault, D. J., Yaxley, R. H., Aveyard, M., Blanchard, A. A., et al. (2005). Perception of motion affects language processing. *Cognition, 94*(3), 79-89.

Kaschak, M. P., Zwaan, R. A., Aveyard, M., & Yaxley, R. H. (2006). Perception of auditory motion affects language processing. *Cognitive Science, 30*, 733–744.

Kelleher, J., & Costello, F. J. (2005). Cognitive Representations of Projective Prepositions. In V. a. V. Kordoni, A. (Ed.), *Proceedings of the 2nd ACL-Sigsem Workshop on The Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications*. Colchester. U.K.: ACL-Sigsem.

Kemmerer, D. (2005). The spatial and temporal meanings of English prepositions can be independently impaired. *Neuropsychologia, 43*(5), 797-806.

Kemmerer, D. (2006). The semantics of space: Integrating linguistic typology and cognitive neuroscience. *Neuropsychologia, 44*(9), 1607-1621.

Kihlstrom, J. F. (1984). Conscious, Subconscious, Unconscious: A Cognitive Perspective. In K. S. Bowers & D. Meichenbaum (Eds.), *The Unconscious Reconsidered* (pp. 149-211). New York: John Wiley & Sons.

Kimmel, M. (2005). Culture regained: Situated and compound image schemas. In B. Hampe & J. E. Grady (Eds.), *From perception to meaning: image schemas in cognitive linguistics* (pp. 285-311). Berlin; New York: Mouton de Gruyter.

King, L., & Clark, J. M. (2002). Intuition and the development of expertise in surgical ward and intensive care nurses. *Journal of Advanced Nursing, 37*(4), 322-329.

Klein, G. A. (1993). A recognition-primed decision (RPD) model of rapid decision making. In G. A. Klein, J. Orasamu, R. Calderwood & C. E. Zsambok (Eds.), *Decision making in action: Models and methods* (pp. 138-147). Norwood, NJ: Ablex.

Klein, G. A. (1998). *Sources of power: How people make decisions*. Cambridge, MA: MIT Press.

Koffka, K. (1935). *Principles of gestalt psychology*. New York,: Harcourt, Brace and Company.

Kolodner, J. (1993). *Case-based reasoning*. San Mateo: Morgan Kaufmann.

Kövecses, Z. (2005). *Metaphor in culture: Universality and variation*. Cambridge; New York: Cambridge University Press.

Krug, S. (2006). *Don't make me think! A common sense approach to Web usability* (2nd ed.). Berkeley, Calif: New Riders.
Krzeszowski, T. P. (1993). The axiological parameter in preconceptional image schemata. In R. A. Geiger & B. Rudzka-Ostyn (Eds.), *Conceptualizations and Mental Processing in Language* (pp. 307-329). Berlin, New York: Mouton de Gruyter.
Krzeszowski, T. P. (1997). *Angels and devils in hell: Elements of axiology in semantics*. Warszawa: Energeia.
Kuhn, W., & Frank, A. U. (1991). A Formalization of Metaphors and Image-Schemas in User Interfaces. In D. M. Mark & A. U. Frank (Eds.), *Cognitive and Linguistic Aspects of Geographic Space* (pp. 419-434). Dordrecht, NL: Kluwer Academic.
Kuijk, J. van (2007). *More iDrive Reviews; the Evolution of a Bad Idea.* Retrieved January 18, 2008, from the product usability weblog, http://www.uselog.com/2007/12/more-bmw-idrive-reviews-evolution-of.html.
Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.
Lakoff, G. (1990). The invariance hypothesis: Is abstract reason based on image-schemas? *Cognitive Linguistics, 1*, 39-74.
Lakoff, G., Espenson, J, Goldberg, A., & Schwartz, A. (1991). *Master metaphor list* (2nd ed.). Retrieved July 13, 2009, from http://araw.mede.uic.edu/~alansz/metaphor/METAPHORLIST.pdf
Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.
Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to Western thought*. New York: Basic Books.
Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.
Laug, S. F. (2003). *The 144 Virtues*. Retrieved March 23, 2008, from http://www.silent-light.de/english/e_THE_144_VIRTUES.html
Leslie, A. M. (1982). The perception of causality in infants. *Perception, 11*(2), 173-186.
Levinson, S., & Meira, S. (2003). 'Natural concepts' in the spatial topological domain – Adpositional meanings in crosslinguistic perspective: An exercise in semantic typology. *Language, 79*(3), 485-516.
Loest, K. (2007). *Image-schematische Metaphern in User Interface. Eine experimentelle Studie mit virtuellen Schiebereglern.* Unpublished Diploma Thesis, Technische Universität Berlin.
Loftus, A. M., Nicholls, M. E. R., Mattingley, J. B., & Bradshaw, J. L. (2008). Left to right: Representational biases for numbers and the effect of visuomotor adaptation. *Cognition, 107*, 1048-1058.
Longo, M. R., & Lourenco, S. F. (2007). Spatial attention and the mental number line: Evidence for characteristic biases and compression. *Neuropsychologia, 45*(7), 1400-1407.
Lund, A. (2003). *Massification of the intangible: An investigation into embodied meaning and information visualization.* Doctoral Thesis. Umeå: Umeå universitet.
Lund, A., & Wiberg, M. (2001). Situating events in roamviz: Using spatiotemporal dimensions to visualize sustained and dynamic mobile projects. In S. Bjørnestad, R. E. Moe, A. I. Mørch & A. L. Opdahl (Eds.), *Proceedings of the 24th Information Systems Research Seminar in Scandinavia (IRIS24)* (Vol. 2, pp. 99-114). Bergen, Norway: Department of Information Science; University of Bergen

Madsen, K. H. (1994). A guide to metaphorical design. *Communications of the ACM, 37*(12), 57-62.

Maeda, J. (2006). *The laws of simplicity*. Cambridge, Mass.: MIT Press.

Maglio, P. P., & Matlock, T. (1999). The conceptual structure of information space. In A. Munro, Benyon, D., & Höök, K. (Ed.), *Social navigation of information space* (pp. 155-173). London: Springer.

Mandler, J. M. (1992). How to build a baby: II. Conceptual primitives. *Psychological Review, 99*(4), 587-604.

Mandler, J. M. (2004). *The foundations of mind: Origins of conceptual thought*. Oxford, New York: Oxford University Press.

Mandler, J. M. (2005). How to build a baby: III. Image schemas and the transition to verbal thought. In B. Hampe & J. E. Grady (Eds.), *From perception to meaning: image schemas in cognitive linguistics* (pp. 137-163). Berlin; New York: Mouton de Gruyter.

Marc, R. (1991). Nobody reads documentation. *Communications od the ACM, 34*(7), 19-24.

Mayhew, D. J. (1992). *Principles and guidelines in software user interface design*. Englewood Cliffs, N.J.: Prentice Hall.

Mayhew, D. J. (1999). *The usability engineering lifecycle: A practitioner's handbook for user interface design*. San Francisco, CA: Morgan Kaufmann.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.

McNeill, D. (2005). *Gesture and thought*. Chicago: University of Chicago Press.

Meier, B. P., & Robinson, M. D. (2004). Why the Sunny Side Is Up. *Psychological Science, 15*(4), 243–247.

Meier, B. P., & Robinson, M. D. (2006). Does “feeling down” mean seeing down? Depressive symptoms and vertical selective attention. *Journal of Research in Personality, 40*(4), 451-461.

Meier, B. P., Robinson, M. D., Crawford, L. E., & Ahlvers, W. J. (2007). When" light" and" dark" thoughts become light and dark responses: Affect biases brightness judgments. *Emotion, 7*(2), 366-376.

Michotte, A. É. (1963). *The perception of causality*. London: Methuen.

Microsoft Press (1995). *The Windows interface guidelines for software design*. Redmond, WA: Microsoft Press.

Miller, R. J. (2007). Another Slant on the Oblique Effect in Drawings and Paintings. *Empirical Studies of the Arts, 25*(1), 41-61.

Mohs, C., Hurtienne, J., Israel, J. H., Naumann, A., Kindsmüller, M. C., Meyer, H. A., et al. (2006). IUUI – Intuitive Use of User Interfaces. In T. Bosenick, M. Hassenzahl, M. Müller-Prove & M. Peissner (Eds.), *Usability Professionals 2006* (pp. 130-133). Stuttgart: German Chapter der Usability Professionals' Association.

Mohs, C., Hurtienne, J., Kindsmüller, M. C., Israel, J. H., Meyer, H. A., & the IUUI Reseacrh Group (2006). IUUI–Intuitive Use of User In-terfaces: Auf dem Weg zu einer wissenschaftlichen Basis für das Schlagwort “Intuitivität”. *MMI-Interaktiv*(11), 75-84.

Mohs, C., Hurtienne, J., Scholz, D., & Rötting, M. (2006). Intuitivität: definierbar, beeinflussbar, überprüfbar. In VDI / VDE-Gesellschaft Mess- und Automatisierungstechnik (Eds.), *Useware 2006* (pp. 215-224). Düsseldorf: VDI-Verlag.

Mohs, C., Lewandowitz, L., Patzlaff, C., & Jürgensohn, T. (2007). DisTracked - workload evaluation for users of mobile applications. In M. Rötting, G. Wozny, A. Klostermann & J. Huss (Eds.), *Prospektive Gestaltung von Mensch-Technik-Interaktion* (pp. 447-452). Düsseldorf: VDI.

Montello, D., Fabrikant, S., Ruocco, M., & Middleton, R. (2003). Testing the First Law of Cognitive Geography on Point-Display Spatializations. In W. Kuhn, M. F. Worboys & S. Timpf (Eds.), *Spatial Information Theory. Foundations of Geographic Information Science. International Conference, COSIT 2003, Ittingen, Switzerland, September 24-28, 2003* (pp. 316-331). Berlin / Heidelberg: Springer.

Munnich, E., Landau, B., & Dosher, B. A. (2001). Spatial language and spatial representation: a cross-linguistic comparison. *Cognition, 81*(3), 171-208.

Murphy, G. L. (1996). On metaphoric representation. *Cognition, 60*(2), 173-204.

Naumann, A., & Hurtienne, J. (2010). Benchmarks for intuitive interaction with mobile devices. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services* (pp. 401-402). Lisbon, Portugal: ACM.

Neale, D. C., & Carroll, J. M. (1997). The role of metaphors in user interface design. In M. Helander, T. K. Landauer & P. Prabhu (Eds.), *Handbook of Human-Computer Interaction* (pp. 441-462). Amsterdam: Elsevier Science.

Nekola, J. C., & White, P. S. (1999). The distance decay of similarity in biogeography and ecology. *Journal of Biogeography, 26*, 867–878.

Neumann, C. (2001). Is Metaphor Universal? Cross-Language Evidence from German and Japanese. *Metaphor and Symbol*, *16* (1/2), 123-142.

Nielsen, J. (1989). Executive summary: Coordinating user interfaces for consistency. In J. Nielsen (Ed.), *Coordinating User Interfaces for Consistency* (pp. 1-7). San Diego: Academic Press.

Noddings, N., & Shore, P. J. (1984). *Awakening the inner eye: Intuition in education*. New York: Teachers College, Columbia University.

Norman, D. (2004). Interaction Design for Automobile Interiors. Retrieved 6 Mar, 2008, from http://www.jnd.org/dn.mss/interaction_des.html

Norman, D. A. (1983). Some observations on mental models. In D. Gentner & A. Stevens (Eds.), *Mental Models* (pp. 7-14). Hillsdale, NJ: Erlbaum.

Norman, D. A. (1990). *The design of everyday things*. New York: Doubleday.

Norman, D. A. (2004). Affordances and Design. Retrieved Oct 24, 2007, from http://www.jnd.org/dn.mss/affordances_and_desi.html

Ostendorf, F. (1994). Zur Taxonomie deutscher Dispositionsbegriffe. In W. Hager & M. Hasselhorn (Eds.), *Handbuch deutschsprachiger Wortnormen* (pp. 383-441). Göttingen: Hogrefe.

Pahl, G., Beitz, W., Feldhusen, J., & Grothe, K. H. (2007). *Engineering design: A systematic approach* (3rd ed.). London: Springer.

Paivio, A. (1990). *Mental representations: A dual-coding approach*. New York: Oxford University Press.

Peeters, B. (2001). Does Cognitive Linguistics Live up to its Name? In R. Dirven, B. Hawkins, & E. Sandikcioglu (Eds.), *Language and Ideology: Theoretical cognitive approaches* (Vol. 1, pp. 83-106). Amsterdam: John Benjamins.

Peña Cervel, M. S. (1999). Subsidiary relationships between image-schemas: An approach to the force schema. *Journal of English Studies, 1*, 187-207.

Proctor, R. W., & Vu, K.-P. L. (2006). *Stimulus-response compatibility principles: Data, theory, and application*. Boca Raton, FL: CRC/Taylor & Francis.

Rakova, M. (2002). The philosophy of embodied realism: A high price to pay? *Cognitive Linguistics, 13*(3), 215-244.

Raskin, J. (1994). Viewpoint: Intuitive equals familiar. *Communications of the ACM, 37*(9), 17-18.

Rasmussen, J. (1986). *Information processing and human-machine interaction: An approach to cognitive engineering*. New York: North-Holland.

Raubal, M. (1997). *Structuring wayfinding tasks with image schemata.* Unpublished Master's Thesis, University of Maine, Orono.

Raubal, M., & Worboys, M. (1999). A Formal Model of the Process of Wayfinding in Built Environments. In C. Freksa & D. M. Mark (Eds.), *Spatial Information Theory: Cognitive Computational Foundations of Geographic Information Science: International Conference COSIT'99. Stade, Germany, August 1999* (pp. 381-400). Heidelberg: Springer.

Richardson, D. C., Spivey, M. J., Barsalou, L. W., & McRae, K. (2003). Spatial representations activated during real-time comprehension of verbs. *Cognitive Science, 27*(5), 767-780.

Richardson, D. C., Spivey, M. J., Edelman, S., & Naples, A. D. (2001). Language is spatial: Experimental evidence for image schemas of concrete and abstract verbs. In K. Stenning & J. Moore (Eds.), *Proceedings of the Twenty-third Annual Meeting of the Cognitive Science Society* (pp. 873-878). Mawah, NJ: Lawrence Erlbaum.

Robert, A. (1997). *From contour completion to image schemas: A modern perspective on gestalt psychology.* Technical Report Cog.Sci UCSD-97.02. San Diego: University of California.

Rohrer, T. (2005). Image schemata in the brain. In B. Hampe & J. E. Grady (Eds.), *From perception to meaning: image schemas in cognitive linguistics* (pp. 165-196). Berlin; New York: Mouton de Gruyter.

Rompay, T. van, Hekkert, P., Saakes, D., & Russo, B. (2005). Grounding abstract object characteristics in embodied interactions. *Acta Psychologica, 119*(3), 315-351.

Rowe, D. W., Sibert, J., & Irwin, D. (1998). Heart rate variability: Indicator of user state as an aid to human-computer interaction. In M. E. Atwood, C. M. Karat, A. Lund, J. Coutaz, & J. Karat (Eds.), *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 480-487). Los Angeles, CA: ACM Press/Addison-Wesley.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge, Mass.: MIT Press.

Rumelhart, D. E., & Ortony, A. (1977). The Representation of Knowledge in Memory. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the Acquisition of Knowledge* (pp. 99-134). Hillsdale, NJ: Lawrence Erlbaum.

Rusconi, E., Kwan, B., Giordano, B. L., Umiltà, C., & Butterworth, B. (2006). Spatial representation of pitch height: The SMARC effect. *Cognition, 99*(2), 113-129.

Schank, R. C. (1982). *Dynamic memory: A theory of reminding and learning in computers and people*. New York: Cambridge University Press.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals and understanding: An inquiry into human knowledge structures*. Hillsdale, NJ: Lawrence Erlbaum.

Schmidtke, H., & Rühmann, H. (1981). Betriebsmittelgestaltung. In H. Schmidtke (Ed.), *Lehrbuch der Ergonomie* (pp. 422-460). München: Hanser.

Schmidt-Kretschmer, M., & Blessing, L. (2006). Strategic aspects of design methodologies: understood or underrated? In D. Marjanovic (Ed.), *Proceedings of the 9th International Design Conference "Design 2006"* (pp. 125-130). Dubrovnik: National University Library.

Schmitt, R. (2001). Metaphern in der Psychologie - eine Skizze. *Journal für Psychologie, 9*(4), 3-15.

Schneider, W., & Chein, J. M. (2003). Controlled & automatic processing: Behavior, theory, and biological mechanisms. *Cognitive Science, 27*(3), 525-559.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*(1), 1-66.

Scholz, D. (2006). *Intuitivität von Mensch-Maschine-Systemen aus Benutzersicht*. Unpublished Diplom Thesis. Berlin: Technische Universität.

Schubert, T. W. (2005). Your highness: Vertical positions as perceptual symbols of power. *Journal of Personality and Social Psychology, 89*(1), 1-21.

Schwibbe, M., Räder, K., Schwibbe, G., Borchardt, M., & Geiken-Pophankem, G. (1994). Zum emotionalen Gehalt von Substantiven, Adjektiven und Verben. In W. Hager & M. Hasselhorn (Eds.), *Handbuch deutschsprachiger Wortnormen* (pp. 272-278). Göttingen: Hogrefe.

Segerer, G. (2008). Closed adjective classes and primary adjectives in African Languages. Retrieved July 13, 2009, from Hyper Article en Ligne - Sciences de l'Homme et de la Société, http://halshs.archives-ouvertes.fr/halshs-00255943/en/

Seth, A., Dienes, Z., Cleeremans, A., Overgaard, M., & Pessoa, L. (2008). Measuring Consciousness: Relating Behavioural and Neurophysiological Approaches. *Trends in Cognitive Sciences, 12*(8), 314-321.

Shepard, R. N. (1974). Representation of structure in similarity data: Problems and prospects. *Psychometrika*, *39*(4), 373-421.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*(2), 127-190.

Sim, J., & Wright, C. C. (2005). The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements. *Physical Therapy, 85*(3), 257.

Spool, J. M. (2005). *What makes a design seem 'intuitive'?* Retrieved Oct 15, 2007, from http://www.uie.com/articles/design_intuitive/

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science, 12*(2), 153-156.

Steen, G. (1999). From linguistic to conceptual metaphor in five steps. In R. W. Gibbs & G. J. Steen (Eds.), *Metaphor in Cognitive Linguistics. Selected papers from the Fifth International Cognitive Linguistics Conference Amsterdam, July 1997* (pp. 57-77). Amsterdam, Philadelphia: John Benjamins.

Stefanowitsch, A. (2006). Words and their metaphors: A corpus-based approach. In A. Stefanowitsch & S. T. Gries (Eds.), *Corpus-based Approaches to Metaphor and Metonymy* (pp. 61-105). Berlin and New York: Mouton de Gruyter.

Sweetser, E. (1990). *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure*. Cambridge, New York: Cambridge University Press.

Sweetser, E. (1998). Regular metaphoricity in gesture: Bodily-based models of speech interaction. In *Actes du 16 e Congrès International des Linguistes.* CD-ROM. Oxford: Elsevier.

Talmy, L. (1983). How language structures space. In H. Pick & L. Acredelo (Eds.), *Spatial orientation: Theory, research, and application* (pp. 225-282). New York: Plenum Press.

Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science, 12*(1), 49-100.

Talmy, L. (2000). *Toward a cognitive semantics*. Cambridge, Mass.: MIT Press.

Talmy, L. (2005). The fundamental system of spatial schemas in language. In B. Hampe & J. E. Grady (Eds.), *From perception to meaning: image schemas in cognitive linguistics* (pp. 199-234). Berlin; New York: Mouton de Gruyter.

Thornton, I. M., & Hubbard, T. L. (2002). Representational momentum: New findings, new directions. *Visual Cognition, 9*(1/2), 1-7.

Tidwell, J. (2006). *Designing interfaces*. Beijing; Sebastopol, CA: O'Reilly.

Traugott, E. C., & Dasher, R. B. (2002). *Regularity in Semantic Change*. Cambridge: Cambridge University Press.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(4), 327-352.

Valenzuela, J., & Soriano, C. (2005). Cognitive Metaphor and Empirical Methods. *Barcelona English Language and Literature Studies*(14), 1-19.

Waller, D., Lippa, Y., & Richardson, A. (2008). Isolating observer-based reference directions in human spatial memory: Head, body, and the self-to-array axis. *Cognition, 106*(1), 157-183.

Weber, K. (2007). *Benutzerorientierte Gestaltung interaktiver Systeme mit Image Schemata am Beispiel SAP R/3.* Unpublished Diplom Thesis, Fachhochschule für Technik und Wirtschaft, Berlin.

Weger, U. W., Meier, B. P., Robinson, M. D., & Inhoff, A. W. (2007). Things are sounding up: Affective influences on auditory tone perception. *Psychonomic Bulletin & Review, 14*(3), 517-521.

Werner, H., & Kaplan, B. (1963). *Symbol formation: An organismic-developmental approach to language and the expression of thought*. New York,: Wiley.

Wickens, C. D., & Carswell, C. M. (1995). The Proximity Compatibility Principle: Its Psychological Foundation and Relevance to Display Design. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 37*, 473-494.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Wierzbicka, A. (1996). *Semantics: Primes and universals*. Oxford; New York: Oxford University Press.

Wilkie, K., Holland, S., & Mulholland, P. (2009). Evaluating musical software using conceptual metaphors. In *Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology* (pp. 232-237). Cambridge, United Kingdom: British Computer Society.

Williams, L. E., Huang, J. Y., & Bargh, J. A. (2009). The scaffolded mind: Higher mental processes are grounded in early experience of the physical world. *European Journal of Social Psychology*, *39*(7), 1257-1267.

Winkielman, P., & Cacioppo, J. T. (2001). Mind at ease puts a smile on the face: Psychophysiological evidence that processing facilitation leads to positive affect. *Journal of Personality and Social Psychology, 81*, 989–1000.

Witt, J. K., Proffitt, D. R., & Epstein, W. (2005). Tool use affects perceived distance, but only when you intend to use it. *Journal of Experimental Psychology: Human Perception and Performance, 31*(5), 880-888.

Wood, L. E. (1998). *User interface design: Bridging the gap from user requirements to design*. Boca Raton: CRC Press.

Yahoo (2008). *Yahoo Developer Network: Design Pattern Library*. Retrieved May 21, 2008, from http://developer.yahoo.com/ypatterns/

Zijlstra, R., & van Doorn, L. (1985). *The construction of a scale to measure subjective effort* Technical Report. Delft, Netherlands: Delft University of Technology, Department of Philosophy and Social Sciences.

Ziori, E., & Dienes, Z. (2006). Subjective measures of unconscious knowledge of concepts. *Mind & Society, 5*(1), 105-122.

Zlatev, J. (2005). What's in a schema? Bodily mimesis and the grounding of language. In B. Hampe & J. E. Grady (Eds.), *From perception to meaning: image schemas in cognitive linguistics* (pp. 313-342). Berlin; New York: Mouton de Gruyter.

Zühlke, D. (2004). *Useware-engineering für technische Systeme*. Berlin: Springer.

Zwaan, R. A., Madden, C. J., Yaxley, R. H., & Aveyard, M. E. (2004). Moving words: Dynamic representations in language comprehension. *Cognitive Science, 28*(4), 611-619.

Zwaan, R. A., & Taylor, L. J. (2006). Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology: General, 135*(1), 1-11.

Zwaan, R. A., & Yaxley, R. H. (2003). Spatial iconicity affects semantic-relatedness judgments. *Psychonomic Bulletin & Review, 10*(4), 954-958.

Zwicky, F. (1969). *Discovery, Invention, Research - Through the Morphological Approach*. Toronto: The Macmillian Company.

A Appendix

The appendix contains additional material used in the experiments (e.g. instruction sheets), detailed results of pre-tests and of statistical procedures of the main experiments. Many of the following tables contain the detailed results of the GLM-repeated-measures procedure and the single comparisons using t-tests for repeated measurements that were conducted in the experiments. When the Mauchly test of sphericity was significant the degrees of freedom were corrected after Greenhouse-Geisser. Effect sizes are reported using partial η^2.

Significances (*p*-values) of post-hoc single comparison tests are reported uncorrected (p) and corrected (p_{corr}) for multiple comparisons (Bonferroni-Holm corrections were applied to achieve a family-wise error rate of $\alpha = .05$). Effect sizes are reported using Cohen's *d*. Asterisks indicating significance refer to corrected *p*-values (if applicable). All *p*-values of single comparisons are reported one-tailed, $* p < .05$, $** p < .01$, $*** p < .001$. Note that reported *p*-values are rounded, but the number of asterisks refers to the not rounded *p*-value.

A.1 Experiment 1: Virtual Push Buttons

A.1.1 Instructions (in German)

Note. Half of the participants received the instruction as presented here (starting with the vertical condition first). The instructions for participants starting with the horizontal version were analogous (not presented here).

Part One (Vertical Condition)

Liebe Versuchsteilnehmerin, lieber Versuchsteilnehmer,

Im folgenden Versuch geht es um die Gestaltung eines Hotel-Informationssystems.

Mit dem System ist es möglich, Informationen z. B. zur Freundlichkeit und Kompetenz des Personals sowie zur Lage, zur Ausstattung und zur Kategorie des jeweiligen Hotels einzugeben.

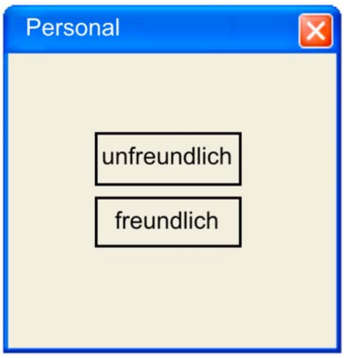
Der Versuch besteht aus zwei Teilen. In beiden Teilen werden Sie aufgefordert, Informationen über 20 Hotels in verschiedenen Städten einzugeben. Dabei geht es nicht um Ihre persönliche Leistung, sondern um die Überprüfung des Systems.

Ihre Daten werden von uns selbstverständlich anonym verarbeitet.

Der Ablauf ist folgendermaßen:

Der Computer präsentiert Ihnen eine kurze Information zum Hotel, z. B. „Das Personal ist freundlich."

Zwei Sekunden später erscheint eine Eingabebox mit zwei Knöpfen, z. B.



Sie machen die jeweils zutreffende Eingabe (hier „freundlich“) über das kleine Eingabegerät.

Benutzen Sie hierfür die obere Taste („8“/“ ↑“), wenn Sie den oberen Knopf betätigen wollen. Benutzen Sie die untere Taste („2“/“ ↓“), wenn Sie den unteren Knopf betätigen wollen.

In unserem Beispiel drücken Sie also die untere Taste („2“/“ ↓“).

Wichtig: Bitte verwenden Sie für die Eingabe zwei Finger!

Zu jedem Hotel werden Ihnen zehn verschiedene Informationen präsentiert.

Nachdem Sie jeweils alle Daten für ein Hotel eingegeben haben, wird das nächste Hotel vorgestellt. An dieser Stelle haben Sie Gelegenheit für eine kurze Pause, denn es geht erst weiter, wenn Sie die ENTER-Taste auf dem Eingabegerät drücken.

Noch ein Hinweis:

Bitte achten Sie darauf, dass Sie möglichst schnell und fehlerfrei arbeiten.

Anhand der Eingaben für das erste Hotel können Sie den Ablauf einmal üben.

Wenn Sie noch Fragen haben, wenden Sie sich bitte an den/die Versuchsleiter/in.

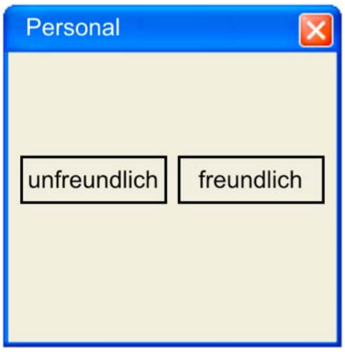
Part Two (Horizontal Condition)

Im nun folgenden Teil des Versuchs werden Ihnen noch einmal Informationen für zwanzig Hotels in verschiedenen Städten dargeboten.

Der Ablauf ist genau so wie in der vorigen Phase. Statt einer senkrechten Anordnung der Knöpfe präsentiert Ihnen der Computer nun die Knöpfe in waagerechter Anordnung an.

Zunächst sehen Sie wieder eine kurze Information zum Hotel, z. B. „Das Personal ist freundlich.“

Zwei Sekunden später erscheint eine Eingabebox, z. B.



Benutzen Sie die linke Taste („4“/“←“), wenn Sie den linken Knopf betätigen wollen und die rechte Taste („6“/“→“), wenn Sie den rechten Knopf betätigen wollen.
In unserem Beispiel drücken Sie also die rechte Taste („6“/“→“).

Wichtig: Bitte verwenden Sie auch hier zwei Finger zur Eingabe!

Zu jedem Hotel werden Ihnen wieder zehn verschiedene Informationen präsentiert.

Bitte achten Sie darauf, dass Sie möglichst schnell und fehlerfrei arbeiten.

Anhand der Eingaben für das erste Hotel können Sie den Ablauf einmal üben.

Wenn Sie noch Fragen haben, wenden Sie sich bitte an den/die Versuchsleiter/in.

A.1.2 Vertical Buttons

Response Times

Table A.1. *Results of the Two-way GLM Repeated Measures Analysis*

Effect	*df*	*F*	Partial η^2	*p*
Mapping***	1,39	33.16	.46	.000
Target domain***	4,156	82.42	.68	.000
Target domain x Mapping***	4,156	9.50	.20	.000

Table A.2. *Results of Single Comparisons between Congruent and Incongruent Mappings Using Paired t-tests*

	Congruent Mapping		Incongruent Mapping						
Target domain	*M*	*SD*	*M*	*SD*	*d*	*t*	*df*	*p*	p_{corr}
Virtue***	733.62	94.00	774.74	110.60	.40	-4.16	39	.000	.000
Quality*	645.41	79.70	671.71	101.53	.29	-2.58	39	.007	.021
Quantity	628.37	83.06	631.27	89.68	.03	-0.42	39	.337	.674
Status	695.39	100.71	694.56	97.31	.01	0.11	39	.457	.457
UP-DOWN***	670.14	98.87	737.86	119.05	.62	-5.61	39	.000	.000

Suitability Judgements

Table A.3. *Results of the Two-way GLM Repeated Measures Analysis*

Effect	*df*	*F*	partial η^2	*p*
Mapping***	1,39	181.51	.82	.000
Target domain**	4,156	4.21	.10	.003
Target domain x Mapping***	2.40,93.66	38.35	.50	.000

Table A.4. *Results of Single Comparisons between Congruent and Incongruent Mappings Using Paired t-tests*

	Congruent Mapping		Incongruent Mapping						
Target domain	*M*	*SD*	*M*	*SD*	*d*	*t*	*df*	*p*	p_{corr}
Virtue***	1.08	0.83	-1.00	0.87	2.45	10.84	39	.000	.000
Quality***	1.26	0.57	-0.99	0.82	3.17	14.71	39	.000	.000
Quantity*	0.40	1.18	-0.15	0.99	.51	1.86	39	.035	.035
Status**	0.64	0.78	0.06	0.91	.68	3.41	39	.001	.002
UP-DOWN***	1.81	0.34	-1.27	1.01	4.07	17.05	39	.000	.000

A.1.3 Horizontal Buttons

Response Times

Table A.5: *Results of the Two-way GLM Repeated Measures Analysis*

Effect	*df*	*F*	Partial η^2	*p*
Mapping	1,39	0.02	.00	.883
Target domain***	3.22,127.67	54.30	.58	.000
Target domain x Mapping***	4,156	5.76	.13	.000

Table A.6. *Results of Single Comparisons between Congruent and Incongruent Mappings Using Paired t-tests*

	Congruent Mapping		Incongruent Mapping						
Target domain	*M*	*SD*	*M*	*SD*	*d*	*t*	*df*	*p*	p_{corr}
Virtue	722.78	107.01	726.77	98.79	.04	0.45	39	.328	.328
Quality*	652.36	88.92	626.87	73.50	.31	-2.55	39	.008	.032
Quantity	613.01	78.11	616.69	88.53	.04	0.60	39	.275	.550
Status	672.22	93.88	662.51	84.34	.11	-1.39	39	.086	,258
UP-DOWN*	651.27	67.40	681.69	91.88	.38	2.94	39	.003	.015

Suitability Judgments

Table A.7: *Results of the Two-way GLM Repeated Measures Analysis*

Effect	*df*	*F*	Partial η^2	*p*
Mapping	1,39	0.72	.02	.402
Target domain*	3.25,126.59	3.14	.07	.025
Target domain x Mapping***	2.38,92.66	23.89	.83	.000

Table A.8. *Results of Single Comparisons between Congruent and Incongruent Mappings Using Paired t-tests*

	Congruent Mapping		Incongruent Mapping						
Target domain	*M*	*SD*	*M*	*SD*	*d*	*t*	*df*	*p*	p_{corr}
Virtue	-0.01	1.24	0.54	1.38	.42	1.57	39	.062	.125
Quality*	-0.21	1.25	0.71	1.32	.72	2.57	39	.007	.021
Quantity***	0.82	1.08	-0.69	0.93	1.50	-5.97	39	.000	.000
Status***	0.69	0.84	-0.08	0.89	.89	-4.37	39	.000	.000
UP-DOWN	0.12	0.95	0.04	1.00	.08	-0.38	39	.354	.354

A.1.4 Importance of Hotel Characteristics

The following table lists the average importance participants gave to different characteristics when evaluating a hotel.

Table A.9. *Importance of Hotel Characteristics*

Target domain	How important are the following criteria for you when evaluating a hotel? (1 – not at all important … 5 – very important)	*M*	*SD*
Virtue	… whether the staff is friendly or unfriendly	4.28	0.82
	… whether the staff is competent or incompetent	4.28	0.64
Quality	… whether the breakfast buffet is good or bad	4.23	0.95
	… whether the rail connections are good or bad	3.60	0.90
Quantity	… whether the hotel is booked by 90% or 70%	1.90	0.93
	… whether the parking garage has 100 or 30 lots	1.68	0.80
Status	… whether the hotel is a luxury or standard hotel	3.18	1.20
	… whether the hotel is in the city centre or in the suburbs	4.08	0.89
Physical UP-DOWN	… whether the hotel bar is up or down	1.58	0.64
	… whether the meeting rooms are up or down	1.53	0.75

Note. Questionnaire items are translated from German.

A.1.5 Word Lengths and Frequencies

The following table lists the length (number of letters) and frequencies of the words used as button labels in experiment one. Word frequencies were determined using the database *Wortschatz Universität Leipzig* (http://wortschatz.uni-leipzig.de). Frequencies are reported in terms of frequency classes. A frequency class of 11, for example, means that the German definite article *der* is 2^{11} times more frequent than the word in question. The higher the frequency class is the less frequent the word is in the linguistic corpus of *Wortschatz Universität Leipzig.*

The results indicate that differences in word lengths were not significant, Wilcoxon test $Z = -1.85$, n.s. Differences in word frequencies reached significance, $Z = -2.12$, $p < .05$. Note that the labels are reported in German (as they were used in the study).

Table A.10. *Word lengths (Number of Letters) and Word frequencies (Frequency Classes) of the Button Labels in Experiment 1*

Target domain	Item #	Label 1			Label 2		
		Word	Length	Frequ.	Word	Length	Frequ.
Virtue	1	freundlich	10	11	unfreundlich	12	15
	2	kompetent	9	13	inkompetent	11	16
Quality	3	gut	3	6	schlecht	8	9
	4	gut	3	6	schlecht	8	9
Quantity	5	90%	3	--	70%	3	--
	6	100	3	--	30	2	--
Status	7	Innenstadt	10	9	Vorstadt	8	13
	8	Luxus	5	12	Standard	8	10
Physical UP-DOWN	9	oben	4	8	unten	5	9
	10	oben	4	8	unten	5	9
Mean (*SD*)			5.40 (3.03)	9.13 (2.64)		7.00 (3.23)	11.25 (2.96)

A.2 Experiment 2: Virtual Sliders

A.2.1 Instructions (in German)

Note. A quarter of the participants received the instruction as presented here (starting with the vertical condition first and showing an example that is metaphor-incongruent). Another quarter of the participants also received the vertical condition first, but with an example that was metaphor-congruent (not presented here). The instructions for the other half of participants starting with the horizontal version were analogous (not presented here). Each part of the instructions fit on one A4 page.

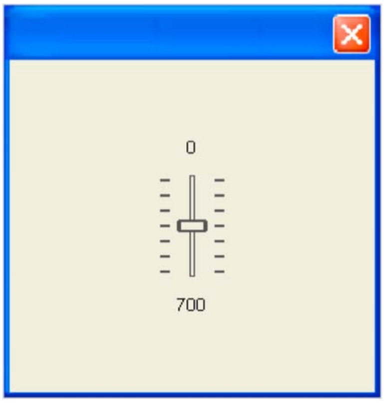
Part One

In dem Experiment geht es um das Einstellen von Schiebereglern. Alle Aufgaben im Teil 1 besitzen die gleiche Struktur. Der Ablauf ist stets folgendermaßen:

1. Um eine Aufgabe zu starten, drücken Sie mit dem rechten Zeigefinger die mittlere Taste „M“ und halten diese gedrückt. Daraufhin erscheint in der Mitte des Bildschirms ein Begriff, der angibt, was auf dem nachfolgenden Schieberegler eingestellt werden soll, zum Beispiel:

   „mehr“.

2. Halten Sie die mittlere Taste weiterhin gedrückt. Der Begriff verschwindet und es erscheint ein auf beiden Seiten beschrifteter Schieberegler:

   Der Regler befindet sich stets auf der mittleren Position. Ihre Aufgabe ist es, anzugeben, in welche Richtung der Regler bewegt werden muss, um das zuvor Gelesene, in unserem Beispiel „mehr“ anzuzeigen.

3. Machen Sie mit dem rechten Zeigefinger die jeweils zutreffende Eingabe über das Tastenfeld:

   - Wenn der Regler nach oben bewegt werden muss, drücken Sie die Taste
     „PFEIL NACH OBEN“.

- Wenn der Regler nach unten bewegt werden muss, drücken Sie die Taste
  „PFEIL NACH UNTEN".

  In dem Beispiel drücken Sie also die Taste „PFEIL NACH UNTEN", da der Regler nach unten bewegt werden muss, um „mehr" anzuzeigen.

4. Nach Ihrer Eingabe beginnt der nächste Durchgang. Sie erhalten wie im Vortest kein Feedback zu Ihrer Eingabe.

Bitte beachten Sie die folgenden Hinweise: Die Beschriftung des Schiebereglers wird sich während des Versuchs häufig ändern!

Sobald Sie die mittlere Taste los lassen, wird die Beschriftung überschrieben. Lassen Sie die mittlere Taste also erst dann los, wenn Sie wissen, in welche Richtung der Regler bewegt werden muss!

Das Experiment beginnt mit einem Übungsblock. Anschließend folgen vier Blöcke mit je 30 Aufgaben. Nach jedem Block gibt es eine kurze Pause, dann geht es erst weiter, wenn Sie die ENTER-Taste auf dem Tastenfeld drücken.

Bitte achten Sie darauf, dass Sie möglichst schnell und fehlerfrei arbeiten!

Wenn Sie noch Fragen haben, wenden Sie sich an die Versuchsleiterin.

Part Two

In der nun folgenden Phase des Experiments werden Ihnen die Schieberegler in waagerechter Anordnung dargeboten. Der Ablauf ist genauso, wie in der vorigen Phase:

1. Um eine Aufgabe zu starten, drücken Sie mit dem rechten Zeigefinger die mittlere Taste „M" und halten diese gedrückt. Daraufhin erscheint in der Mitte des Bildschirms ein Begriff, der angibt, was auf dem nachfolgenden Schieberegler eingestellt werden soll, zum Beispiel:

   „mehr".

2. Halten Sie die mittlere Taste weiterhin gedrückt. Der Begriff verschwindet und es erscheint ein auf beiden Seiten beschrifteter Schieberegler:

Der Regler befindet sich stets auf der mittleren Position. Ihre Aufgabe ist es, anzugeben, in welche Richtung der Regler bewegt werden muss, um das zuvor Gelesene, in unserem Beispiel „mehr“ anzuzeigen.

3. Machen Sie mit dem rechten Zeigefinger die jeweils zutreffende Eingabe über das Tastenfeld:
   - Wenn der Regler nach links bewegt werden muss, drücken Sie die Taste
     „PFEIL NACH LINKS“.
   - Wenn der Regler nach rechts bewegt werden muss, drücken Sie die Taste
     „PFEIL NACH RECHTS“.

   In dem Beispiel drücken Sie also die Taste „PFEIL NACH LINKS“, da der Regler nach links bewegt werden muss, um „mehr“ anzuzeigen.
4. Nach Ihrer Eingabe beginnt der nächste Durchgang. Sie erhalten kein Feedback zu Ihrer Eingabe.

Auch hier gelten wieder die Hinweise: Die Beschriftung des Schiebereglers wird sich während des Versuchs häufig ändern!

Sobald Sie die mittlere Taste los lassen, wird die Beschriftung überschrieben. Lassen Sie die mittlere Taste also erst dann los, wenn Sie wissen, in welche Richtung der Regler bewegt werden muss!

Dieser Teil des Experiments beginnt ebenfalls mit einem Übungsblock. Anschließend folgen vier Blöcke mit je 30 Aufgaben. Nach jedem Block gibt es eine kurze Pause, dann geht es erst weiter, wenn Sie die ENTER-Taste auf dem Tastenfeld drücken.

Bitte achten Sie darauf, dass Sie möglichst schnell und fehlerfrei arbeiten!

Wenn Sie noch Fragen haben, wenden Sie sich an die Versuchsleiterin.

A.2.2 Slider Labels

Target Domain Valence

Table A.11 lists the words used as labels in the slider experiment along with their characteristics for the target domain of valence. Word length, indicated as the number of letters, differed between the two labels of a pair, Wilcoxon test statistic $Z = -3.27$, $p < .01$.

As above, word frequencies were determined using the database *Wortschatz Universität Leipzig* (http://wortschatz.uni-leipzig.de). Differences in word frequencies were significant, Wilcoxon test statistic $Z = -2.77$, $p < .01$.

The next two columns in the table are ratings of how much the adjective is related to quality or quantity (1 – not at all, 7 – very strongly). Relatedness to quality was significantly greater than five, $t(23) = 5.72$, $p < .001$. Relatedness to quantity was significantly lower than three, $t(23) = -5.15$, $p < .001$. The last two columns of Table A.11 contain the semantic differential ratings on the dimensions unpleasant-pleasant and bad-good. Because the adjectives of these dimensions also were used as slider labels, e.g. no bad-good rating is given for the words *good* and *bad* (*gut* and *schlecht* in German). Unpleasant-pleasant and bad-good ratings differed significantly between the two words of a pair of labels, $t(10) = 29.96$, $p < .001$, and $t(10) = 33.45$, $p < .001$, respectively. Note that the labels are reported in German (as they are used in the study).

Target Domain Quantity

Slider labels in the target domain of quantity were 12 pairs of numbers:

1 – 0	10 – 0	100 – 0	3982 – 0
5 – 0	50 – 0	289 – 0	5000 – 0
6 – 0	66 – 0	500 – 0	8000 – 0

Target Domain Physical UP-DOWN

Slider labels in the target domain of physical UP-DOWN are listed along with their word lengths and frequencies in Table A.12. Word lengths, indicated as the number of letters, differed between the two labels of a pair, Wilcoxon test statistic $Z = -2.24$, $p < .05$. Word frequencies, indicated as frequency classes as above, showed no significant differences between the two labels of a pair, $Z = -1.67$, n.s. Note that the labels are reported in German (as they were used in the study).

Table A.11. *Word Lengths (Number of Letters) and Frequencies (Frequency Classes), Relatedness to Quality and Quantity (1 – not at all … 7 – very strong) and Ratings on the Dimensions unpleasant-pleasant and bad-good (1…7) of the Slider Labels*

Label 1							Label 2						
Word	Length	Frequ.	Qual.	Quan.	Pleas.	Good	Word	Length	Frequ.	Qual.	Quan.	Pleas.	Good
angenehm	8	12	5.04	2.57	--	6.49	unangenehm	10	12	4.69	2.29	--	1.73
beliebt	7	11	5.77	2.86	6.09	5.95	unbeliebt	9	14	5.73	2.68	1.73	2.23
ehrlich	7	11	5.83	2.48	6.14	6.18	unehrlich	9	15	5.26	2.26	1.52	1.38
freundlich	10	11	5.95	1.65	6.38	6.33	unfreundlich	12	15	5.71	2.62	1.86	1.75
gütig	5	16	5.45	1.95	6.05	6.10	boshaft	7	16	5.09	2.27	1.91	1.52
gut	3	6	5.53	3.07	6.72	--	schlecht	8	9	5.29	2.53	1.57	--
kompetent	9	13	6.52	3.04	6.05	6.09	inkompetent	11	16	5.78	2.68	2.09	1.81
stabil	6	11	5.92	2.87	5.68	5.63	instabil	8	14	5.29	2.58	1.78	2.00
originell	9	14	5.35	2.65	5.48	5.91	langweilig	10	12	5.22	2.52	2.00	2.13
praktisch	9	10	5.32	3.00	5.95	6.14	unpraktisch	11	16	5.00	2.68	2.24	2.10
tolerant	8	14	5.50	2.58	6.21	6.21	intolerant	10	17	4.83	2.17	1.71	1.78
zufrieden	9	8	5.83	3.33	6.35	6.33	unzufrieden	11	11	5.67	3.04	2.04	2.26
Mean (*SD*)	7.50 (2.02)	11.42 (2.71)	5.67 (0.39)	2.67 (0.48)	6.10 (0.34)	6.12 (0.24)	Mean (*SD*)	9.67 (1.50)	13.92 (2.43)	5.30 (0.36)	2.53 (0.25)	1.86 (0.22)	1.88 (0.29)

Table A.12. *Word Lengths (Number of Letters) and Frequencies (Frequency Classes) for the Slider Labels of the Target Domain of Physical* UP-DOWN

Label 1			Label 2		
Word	Length	Frequ.	Word	Length	Frequ.
darüber	7	7	darunter	8	8
drüber	6	12	drunter	7	14
hoch	4	8	tief	4	9
oben	4	8	unten	5	9
oberhalb	8	12	unterhalb	9	11
zuoberst	8	19	zuunterst	9	20
Mean (*SD*)	6.17 (1.84)	11.00 (4.47)	Mean (*SD*)	7.00 (2.10)	11.83 (4.54)

A.2.3 Vertical Sliders

Response Times

Table A.13. *Results of the Two-way GLM Repeated Measures Analysis*.

Effect	*df*	*F*	Partial η^2	*p*
Mapping***	1,39	42.70	.52	.000
Target domain***	2,78	176.17	.82	.000
Target domain x Mapping**	1.60,62.21	8.43	.18	.001

Table A.14. *Results of Single Comparisons between Congruent and Incongruent Mappings Using Paired t-tests*

	Congruent Mapping		Incongruent Mapping						
Target domain	*M*	*SD*	*M*	*SD*	*d*	*t*	*df*	*p*	p_{corr}
Valence**	935.11	196.85	968.77	209.00	.17	-2.47	39	.009	.009
Quantity***	635.18	107.09	682.69	146.71	.37	-4.60	39	.000	.000
UP-DOWN***	904.57	185.20	1003.71	222.75	.48	-6.35	39	.000	.000

Suitability Judgments

Table A.15. *Results of the Two-way GLM Repeated Measures Analysis*

Effect	*df*	*F*	Partial η^2	*p*
Mapping***	1,39	282.76	.88	.000
Target domain*	1.52,59.07	3.57	.08	.046
Target domain x Mapping***	2,78	14.18	.27	.000

Table A.16. *Results of Single Comparisons between Congruent and Incongruent Mappings Using Paired t-tests*

	Congruent Mapping		Incongruent Mapping						
Target domain	*M*	*SD*	*M*	*SD*	*d*	*t*	*df*	*p*	p_{corr}
Valence***	1.05	0.50	-0.97	0.74	3.20	13.91	39	.000	.000
Quantity***	1.44	0.69	-0.99	1.08	2.68	10.99	39	.000	.000
UP-DOWN***	1.60	0.64	-1.46	0.91	3.89	16.30	39	.000	.000

A.2.4 Horizontal Sliders

Response Times

Table A.17. *Results of the Two-way GLM Repeated Measures Analysis*

Effect	*df*	*F*	Partial η^2	*p*
Mapping	1,39	0.91	.02	.347
Target domain***	2,78	309.39	.89	.000
Target domain x Mapping	2,78	.537	.01	.587

Table A.18. *Results of Single Comparisons between Congruent and Incongruent Mappings Using Paired t-tests*

	Congruent Mapping		Incongruent Mapping						
Target domain	*M*	*SD*	*M*	*SD*	*d*	*t*	*df*	*p*	p_{corr}
Valence	966.95	171.64	952.15	170.48	.09	1.33	39	.095	.286
Quantity	630.20	102.38	633.27	114.37	.03	-0.49	39	.314	.314
UP-DOWN	991.49	168.42	980.99	177.23	.06	0.57	39	.287	.574

Suitability Judgments

Table A.19. *Results of the Two-way GLM Repeated Measures Analysis*

Effect	*df*	*F*	Partial η^2	*p*
Mapping***	1,39	113.72	.75	.000
Target domain***	2,78	28.24	.42	.000
Target domain x Mapping***	2,78	42.61	.52	.000

Table A.20: *Results of Single Comparisons between Congruent and Incongruent Mappings Using Paired t-tests*

	Congruent Mapping		Incongruent Mapping						
Target domain	*M*	*SD*	*M*	*SD*	*d*	*t*	*df*	*p*	p_{corr}
Valence***	1.00	0.58	-0.38	0.92	1.79	7.30	39	.000	.000
Quantity***	1.56	0.39	-0.86	1.10	2.93	13.29	39	.000	.000
UP-DOWN***	0.38	0.75	-1.00	0.73	1.86	8.38	39	.000	.000

A.3 Experiment 3: Single Comparisons

A.3.1 Instructions (in German)

Note. Half of the participants received the instruction as presented here (left key for similar, right key for different). The other half of the participants received instructions with a reversed assignment of response keys (not presented here).

Liebe Versuchsteilnehmerin, lieber Versuchsteilnehmer,

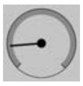
Beim Umgang mit Technik müssen oft Werte verglichen werden, z. B. IST-SOLL-Werte, Parameter verschiedener Maschinen oder Wirtschaftsdaten von Produktionsstandorten. Es gibt dazu zwei Möglichkeiten der Anzeige: numerische Anzeigen und zeigerbasierte Anzeigen. Beide sollen in diesem Experiment untersucht werden. Dabei geht es um Vergleiche der Ähnlichkeit bzw. Verschiedenheit von Anzeigewerten.

Dabei sollen Sie folgende Definitionen von *Ähnlichkeit* verwenden:

Bei den **numerischen Anzeigen** bedeutet Ähnlichkeit ausschließlich die *Übereinstimmung von Kennzahlen*. Zum Beispiel sind die Anzeigen „50“ und „40“ sich ähnlich (kleine Differenz von 10). Die Anzeigen „10“ und „70“ sind voneinander verschieden (große Differenz von 60). Insgesamt sind zehn verschiedene Kennzahlen im Bereich von 10 bis 100 möglich.

Bei den **zeigerbasierten Anzeigen** bedeutet Ähnlichkeit ausschließlich die *Übereinstimmung von Zeigerwinkeln auf einer Skala*. Die Skala ist der dunkelgraue Bereich auf den Zeigerdarstellungen.

Zum Beispiel sind die Anzeigen

 und  sich ähnlich (kleiner Winkelunterschied auf der Skala). Die Anzeigen

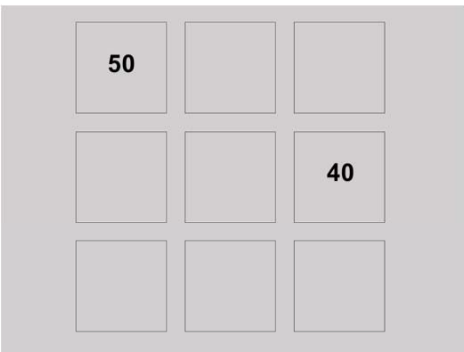
 und  sind voneinander verschieden (großer Winkelunterschied auf der Skala). Insgesamt sind zehn verschiedene Zeigerstellungen im angezeigten Skalenbereich möglich.

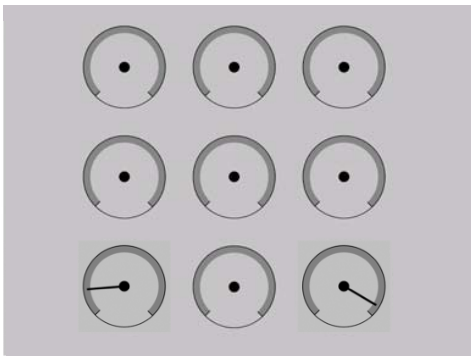
In diesem Experiment sollen Sie entscheiden, ob zwei Anzeigen sich ähnlich oder voneinander verschieden sind. Der Versuch ist in zwei Teile gegliedert. In einem Teil geht es um numerische Anzeigen, im zweiten Teil um zeigerbasierte Anzeigen.

Bei der Bewertung der Ähnlichkeit der Anzeigen gibt es keine richtigen oder falschen Antworten. Wichtig ist nur, dass sie die Ähnlichkeit oder Verschiedenheit der Anzeigen in der oben angegebenen Weise einschätzen.

Der **Ablauf** ist in beiden Teilen gleich:

1. Auf dem Bildschirm erscheinen neun Felder mit Anzeigen, von denen zwei Anzeigen entweder mit Zahlen (im Teil mit numerischen Anzeigen) oder mit Zeigern (im Teil mit zeigerbasierten Anzeigen) gefüllt sind. Diese beiden Anzeigen sollen Sie miteinander vergleichen.





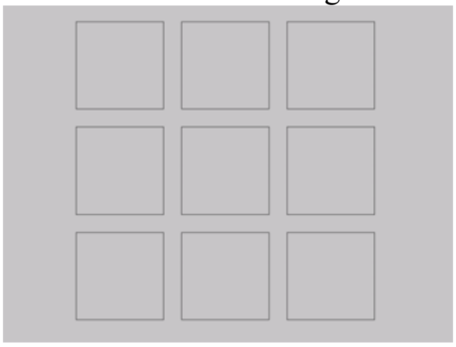
2. Wenn Sie meinen, die beiden Anzeigen sind sich nach der obigen Definition ähnlich, drücken Sie die LINKE rote Taste auf der Tastatur. Wenn Sie meinen, die beiden Anzeigen sind verschieden, drücken Sie die RECHTE rote Taste auf der Tastatur.

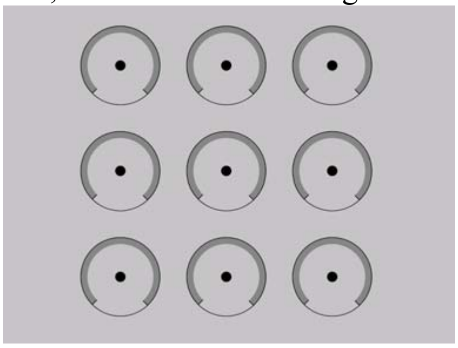
Für die obigen Beispiele bedeutet das demnach:

Die numerischen Anzeigen sind sich ähnlich. Sie drücken also die LINKE rote Taste.

Die zeigerbasierten Anzeigen sind voneinander verschieden. Sie drücken also die RECHTE rote Taste.

3. Nachdem Sie eine rote Taste gedrückt haben, leeren sich alle Anzeigen.





Nach kurzer Zeit werden zwei neue Anzeigenfelder gefüllt, die sie wiederum vergleichen sollen (s. Punkt 1).

Beide Teile bestehen aus je 6 Blöcken. Am Anfang eines Teiles steht jeweils ein Übungsblock, in dem Sie den Ablauf üben können.

Zwischen den Blöcken haben Sie jeweils die Gelegenheit für eine kleine Pause. Mit dem Drücken der grünen Taste können Sie dann mit dem Versuch fortfahren.

Nach dem letzten Block eines jeden Teiles werden wir Sie jeweils bitten, einen Fragebogen auszufüllen.

Wichtige Hinweise:

Bitte benutzen Sie zur Betätigung der LINKEN roten Taste die linke Hand und für die Betätigung der RECHTEN rote Taste die rechte Hand.

Es geht es nicht darum, zu messen, wie gut Sie persönlich abschneiden, sondern welche Eingabevariante am besten funktioniert. Ihre Daten werden selbstverständlich anonym behandelt.

Bitte achten Sie darauf, dass Sie möglichst schnell und akkurat arbeiten.

Wenn Sie noch Fragen haben, wenden Sie sich bitte an die Versuchsleiterin.

A.3.2 Similarity Thresholds

Table A.21. *Results of the Comparisons of Individual Threshold Values Using Paired t-tests*

	NEAR		FAR					
Display type	*M*	*SD*	*M*	*SD*	*d*	*t*	*df*	*p*
Number	26.46	11.75	26.88	11.11	.04	.811	23	.213
Pointer*	27.08	11.41	26.04	10.42	.10	-1.74	23	.048

Note. Data for the pointer display type are numerical values (not degrees).

A.3.3 Error Rates

Table A.22. *Results of the Three-way GLM Repeated Measures Analysis*

Effect	*df*	*F*	Partial η^2	*p*
Display type**	1,23	9.27	.29	.006
Display proximity	1,23	.50	.02	.488
Similarity**	1,23	10.85	.32	.003
Display type x Display proximity*	1,23	6.68	.23	.017
Display type x Similarity	1,23	2.51	.10	.127
Display proximity x Similarity	1,23	.68	.03	.419
Display type x Display proximity x Similarity	1,23	.51	.02	.484

Table A.23. *Results of Single Comparisons between NEAR and FAR Display Proximities Using Paired t-tests*

Display	Values	NEAR		FAR						
		M	*SD*	*M*	*SD*	*d*	*t*	*df*	*p*	p_{corr}
Number	Different*	3.79	4.35	1.48	1.79	.70	2.83	23	.005	.019
	Similar	7.13	5.30	6.47	4.12	.14	0.59	23	.280	.559
Pointer	Different	5.96	4.37	6.73	4.06	.18	-0.83	23	.207	.621
	Similar	7.38	6.05	8.05	5.91	.11	-0.54	23	.296	.296

Note. Error rates are in %.

A.3.4 Response Times

Table A.24. *Results of the Three-way GLM Repeated Measures Analysis*

Effect	*df*	*F*	Partial η^2	*p*
Display type	1,23	.01	.00	.939
Display proximity*	1,23	7.34	.24	.013
Similarity***	1,23	19.42	.46	.000
Display type x Display proximity	1,23	2.41	.10	.134
Display type x Similarity**	1,23	10.19	.31	.004
Display proximity x Similarity	1,23	.20	.01	.658
Display type x Display proximity x Similarity	1,23	1.41	.06	.248

Table A.25. *Results of Single Comparisons between* NEAR *and* FAR *Display Proximities Using Paired t-tests*

Display	Values	NEAR *M*	NEAR *SD*	FAR *M*	FAR *SD*	*d*	*t*	*df*	*p*	p_{corr}
Number	Diff.*	1072.34	166.83	1128.29	199.80	.30	2.52	23	.010	.038
	Sim.*	1045.41	155.46	1087.68	166.70	.26	2.44	23	.011	.034
Pointer	Diff.	1167.09	404.03	1163.94	344.07	.01	-0.13	23	.450	.450
	Sim.	980.15	235.21	1011.23	223.22	.14	1.29	23	.105	.209

A.3.5 Suitability Judgments

Table A.26. *Results of the Three-way GLM Repeated Measures Analysis*

Effect	*df*	*F*	Partial η^2	*p*
Display type	1,23	.08	.00	.784
Display proximity***	1,23	33.41	.59	.000
Similarity***	1,23	38.67	.63	.000
Display type x Display proximity	1,23	1.70	.07	.205
Display type x Similarity**	1,23	11.10	.33	.003
Display proximity x Similarity	1,23	.15	.01	.698
Display type x Display proximity x Similarity	1,23	.00	.00	.994

Table A.27. *Results of Single Comparisons between* NEAR *and* FAR *Display Proximities Using Paired t-tests*

Display	Values	NEAR *M*	NEAR *SD*	FAR *M*	FAR *SD*	*d*	*t*	*df*	*p*	p_{corr}
Number	Different**	0.68	0.65	0.09	0.42	1.08	3.33	23	.001	.001
	Similar***	0.96	0.54	0.33	0.68	1.03	4.52	23	.000	.000
Pointer	Different***	0.39	0.60	-0.02	0.43	.79	3.80	23	.000	.001
	Similar***	1.10	0.45	0.65	0.57	.88	5.46	23	.000	.000

A.4 Experiment 4: Double Comparisons

A.4.1 Instructions (in German)

Note. Half of the participants received the instruction as presented here (starting with the number-display condition first). The instructions for participants starting with the pointer-display version were analogous (not presented here).

Part One

Liebe Versuchsteilnehmerin, lieber Versuchsteilnehmer,

Beim Umgang mit Technik müssen oft Werte verglichen werden, z. B. IST-SOLL-Werte, Parameter verschiedener Maschinen oder Wirtschaftsdaten von Produktionsstandorten. Es gibt dazu zwei Möglichkeiten der Anzeige: Zahlendarstellungen und Zeigerdarstellungen. Beide sollen in diesem Experiment untersucht werden. Im ersten Teil geht es um Zahlendarstellungen.

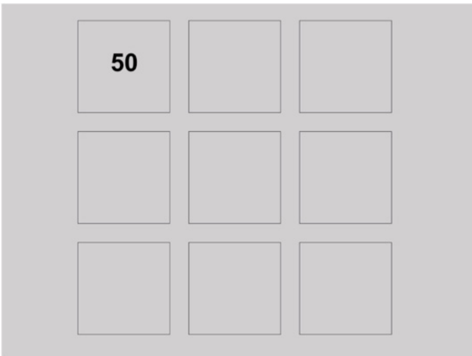
Dabei geht es um Vergleiche der Ähnlichkeit oder Verschiedenheit von Anzeigewerten. Ähnlichkeit heißt dabei Übereinstimmung von Kennzahlen, d.h. ein Wert von 50 ist der 40 ähnlicher (Differenz = 10) als der 70 (Differenz = 20).

In diesem Experiment sollen Sie entscheiden, welche von zwei Anzeigen einer vorgegebenen Anzeige ähnlicher oder von ihr verschiedener ist.

Der Ablauf ist jedes Mal so:

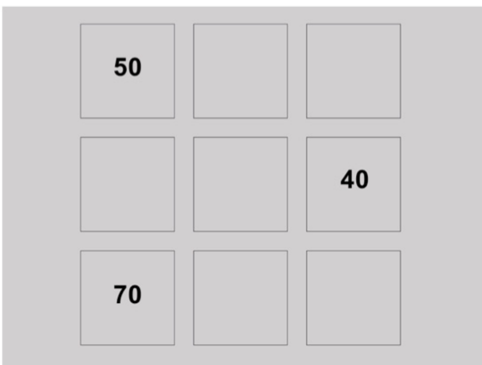
1. Auf dem Bildschirm erscheint entweder die Frage „Welche Anzeigen sind sich ähnlicher?“ oder die Frage „Welche Anzeigen sind verschiedener?“

2. Kurze Zeit danach erscheinen auf dem Bildschirm neun Anzeigen, von denen eine mit einer Zahl gefüllt ist. Dies ist die Anzeige mit der Sie weitere Anzeigen vergleichen sollen.



3. Drücken Sie auf dem Eingabegerät diejenige rote Taste, die der Position der Zahl auf dem Bildschirm entspricht. In unserem Beispiel wird die Zahl 50 in der oberen linken Anzeige angezeigt, die entsprechende Taste auf dem Eingabegerät wäre also oben links zu drücken.

Wenn Sie diese Taste gedrückt haben, erscheinen zwei weitere Zahlen, die Sie mit der ersten Zahl vergleichen sollen. Alle drei Anzeigen sind nur so lange zu sehen, wie Sie die Taste gedrückt halten.



Entscheiden Sie nun, welche der beiden neu erschienenen Zahlen der Aufgabenstellung („Welche Anzeigen sind sich ähnlicher?“/„ Welche Anzeigen sind verschiedener?“) entspricht.

4. Drücken Sie die entsprechende rote Taste auf dem kleinen Eingabegerät.

In unserem Beispiel wäre bei der Aufgabenstellung „Welche Anzeigen sind verschiedener?“ die 70 die richtige Antwort. Da sie unten links steht, drücken Sie die untere linke Taste auf dem Eingabegerät.

Bei der Aufgabenstellung „Welche Anzeigen sind sich ähnlicher?“ wäre dagegen die 40 die richtige Antwort. Da sie in der Mitte rechts steht, drücken Sie die mittlere rechte Taste auf dem Eingabegerät.

Der 1. Teil des Versuches besteht aus 6 Blöcken (3 Blöcke mit der Frage „Welche Anzeigen sind sich ähnlicher?“ und 3 Blöcke mit der Frage „Welche Anzeigen sind verschiedener?“). Hinzu kommt jeweils ein Übungsblock, in den Sie den Ablauf üben können.

Zwischen den Blöcken haben Sie jeweils die Gelegenheit für eine kleine Pause. Mit dem Drücken der grünen Taste können Sie dann mit dem Versuch fortfahren.

Wichtige Hinweise:

Bitte benutzen Sie zur Eingabe nur einen Finger (z.B. nur den Zeigefinger).

Es ist gut, den Finger zwischen den Aufgaben über der mittleren roten Taste „zwischenzuparken". Dann können Sie die anderen Tasten besser erreichen.

Bitte achten Sie darauf, dass Sie möglichst schnell und fehlerfrei arbeiten.

Dabei geht es nicht darum, zu messen, wie gut Sie persönlich abschneiden, sondern welche Eingabevariante am besten funktioniert. Ihre Daten werden selbstverständlich anonym behandelt.

Wenn Sie noch Fragen haben, wenden Sie sich bitte an die Versuchsleiterin.

Part Two

Liebe Versuchsteilnehmerin, lieber Versuchsteilnehmer,

In diesem Teil des Versuches sollen nun Zeigerdarstellungen untersucht werden. Auch hier geht es um Vergleiche der Ähnlichkeit oder Verschiedenheit von Anzeigewerten. Ähnlichkeit heißt dabei Übereinstimmung von Zeigerwinkeln auf einer Skala (der dunkelgraue Bereich auf den folgenden Zeigerdarstellungen). Das heißt,

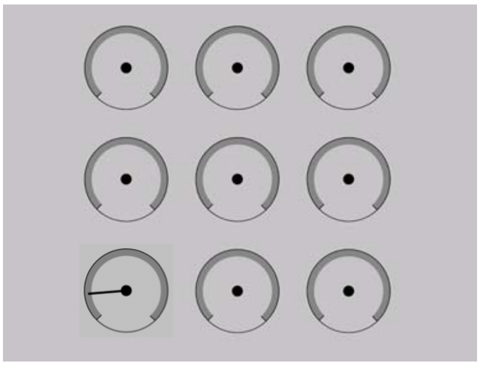
 und  sind sich ähnlicher (kleiner Winkelunterschied auf der Skala) als  und  (großer Winkelunterschied auf der Skala).

Auch in diesem Teil des Versuches sollen Sie entscheiden, welche von zwei Anzeigen einer vorgegebenen Anzeige ähnlicher oder von ihr verschiedener ist.

Der Ablauf ist ähnlich wie im 1. Teil:

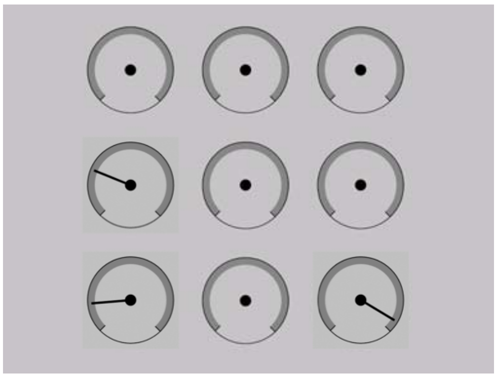
1. Auf dem Bildschirm erscheint entweder die Frage „Welche Anzeigen sind sich ähnlicher?“ oder die Frage „Welche Anzeigen sind verschiedener?“

2. Kurze Zeit danach erscheinen auf dem Bildschirm neun Anzeigen, von denen eine mit einem Zeiger versehen ist. Dies ist die Anzeige mit der Sie weitere Anzeigen vergleichen sollen.



3. Drücken Sie auf dem Eingabegerät diejenige rote Taste, die der Position der Anzeige auf dem Bildschirm entspricht. In unserem Beispiel erscheint die Anzeige im unteren linken Feld, die entsprechende Taste auf dem Eingabegerät wäre also unten links zu drücken.

Wenn Sie diese Taste gedrückt haben, erscheinen zwei weitere Anzeigen, die Sie mit der ersten Anzeige vergleichen sollen. Alle drei Anzeigen sind nur so lange zu sehen, wie Sie die Taste gedrückt halten.



Entscheiden Sie nun, welche der beiden neu erschienenen Anzeigen der Aufgabenstellung („Welche Anzeigen sind sich ähnlicher?“/„ Welche Anzeigen sind verschiedener?“) entspricht.

4. Drücken Sie die entsprechende rote Taste auf dem kleinen Eingabegerät.

In unserem Beispiel wäre bei der Aufgabenstellung „Welche Anzeigen sind verschiedener?“ die Anzeige unten rechts die richtige Antwort. Drücken Sie also die untere rechte Taste auf dem Eingabegerät.

Bei der Aufgabenstellung „Welche Anzeigen sind sich ähnlicher?“ wäre die Anzeige Mitte links die richtige Antwort. Drücken Sie also die mittlere linke Taste auf dem Eingabegerät.

Der 2. Teil des Versuches besteht ebenfalls aus 6 Blöcken (3 Blöcke mit der Frage „Welche Anzeigen sind sich ähnlicher?“ und 3 Blöcke mit der Frage „Welche Anzeigen sind verschiedener?“). Hinzu kommt jeweils ein Übungsblock, in den Sie den Ablauf üben können.

Wichtige Hinweise:

Bitte benutzen Sie zur Eingabe nur einen Finger (z.B. nur den Zeigefinger).

Bitte achten Sie darauf, dass Sie möglichst schnell und fehlerfrei arbeiten.

Wenn Sie noch Fragen haben, wenden Sie sich bitte an die Versuchsleiterin.

A.4.2 Manipulation Check

Table A.28. *Comparisons of the Performance Data between Experiments 3 and 4 Using t-tests for Independent Samples*

Display	Performance variable	Experiment 3		Experiment 4					
		M	*SD*	*M*	*SD*	*d*	*t*	*df*	*p*
Number	Resp. time	1094.95	171.72	1162.71	324.92	.26	.90	46	.371
	Accuracy*	0.94	0.04	0.97	0.03	.85	2.17	46	.035
Pointer	Resp. time***	1102.24	298.74	1967.16	329.62	2.75	9.53	46	.000
	Accuracy*	0.92	0.04	0.88	0.11	.48	-1.73	46	.048

A.4.3 Error Rates

Table A.29. *Results of the Three-way GLM Repeated Measures Analysis*

Effect	*df*	*F*	Partial η^2	*p*
Display type***	1,23	34.10	.60	.000
Display arrangement***	1,23	22.86	.50	.000
Task	1,23	1.92	.08	.180
Display type x Display arrangement**	1,23	10.83	.32	.003
Display type x Task	1,23	1.01	.04	.325
Display arrangement x Task	1,23	.441	.02	.513
Display type x Display arrangement x Task	1,23	.02	.00	.894

Table A.30. *Single Comparisons between simNEAR and diffNEAR Display Arrangements Using Paired t-tests*

Display	Task	simNEAR *M*	simNEAR *SD*	diffNEAR *M*	diffNEAR *SD*	*d*	*t*	*df*	*p*	p_{corr}
Number	Different	2.78	4.30	3.29	3.87	.12	-0.53	23	.299	.299
Number	Similar*	0.00	0.00	1.91	3.50	.77	-2.68	23	.007	.020
Pointer	Different**	6.64	6.10	12.68	10.98	.68	-3.57	23	.001	.003
Pointer	Similar*	5.87	4.60	12.83	12.18	.76	-2.60	23	.008	.016

Note. Error rates are in %.

A.4.4 Response Times

Table A.31. *Results of the Three-way GLM Repeated Measures Analysis*

Effect	*df*	*F*	Partial η^2	*p*
Display type***	1,23	210.52	.90	.000
Display arrangement ***	1,23	28.24	.55	.000
Task***	1,23	67.12	.75	.000
Display type x Display arrangement *	1,23	6.24	.21	.020
Display type x Task	1,23	2.08	.08	.163
Display arrangement x Task	1,23	1.06	.04	.313
Display type x Display arrangement x Task	1,23	.72	.03	.404

Table A.32. *Single Comparisons between simNEAR and diffNEAR Display Arrangements Using Paired t-tests*

Display	Task	simNEAR		diffNEAR						
		M	*SD*	*M*	*SD*	*d*	*t*	*df*	*p*	p_{corr}
Number	Diff.*	1193.34	287.84	1295.79	312.40	.34	1.96	23	.031	.031
	Sim.***	904.20	145.12	1005.97	202.32	.58	4.84	23	.000	.000
Pointer	Diff.*	1941.36	348.62	2111.54	389.81	.46	2.30	23	.015	.031
	Sim.***	1696.67	365.21	1977.43	427.29	.71	4.10	23	.000	.001

A.4.5 Suitability Judgments

Table A.33. *Results of the Three-way GLM Repeated Measures Analysis*

Effect	*df*	*F*	Partial η^2	*p*
Display type*	1,23	4.33	.16	.049
Display arrangement ***	1,23	14.25	.38	.001
Task	1,23	.50	.02	.487
Display type x Display arrangement	1,23	1.82	.07	.191
Display type x Task	1,23	.01	.00	.942
Display arrangement x Task**	1,23	8.55	.27	.008
Display type x Display arrangement x Task*	1,23	6.72	.23	.016

Table A.34. *Single Comparisons between simNEAR and diffNEAR Display Arrangements Using Paired t-tests*

Display	Task	simNEAR		diffNEAR						
		M	*SD*	*M*	*SD*	*d*	*t*	*df*	*p*	p_{corr}
Number	Different	0.28	0.72	0.49	0.63	.31	1.01	23	.161	.322
	Similar***	0.89	0.65	-0.05	0.71	1.38	-4.54	23	.000	.000
Pointer	Different	0.28	0.47	0.19	0.52	.18	-0.74	23	.233	.233
	Similar*	0.47	0.47	0.09	0.55	.74	-2.58	23	.008	.025

A.4.6 Additional Analyses of the Filler Items

Error Rates

Table A.35. *Results of the Three-way GLM Repeated Measures Analysis*

Effect	*df*	*F*	Partial η^2	*p*
Display type***	1,23	14.34	.38	.001
Display position	1,23	2.56	.10	.123
Task*	1,23	7.75	.25	.011
Display type x Display position	1,23	.65	.03	.428
Display type x Task	1,23	.00	.00	.999
Display position x Task	1,23	.04	.00	.844
Display type x Display position x Task	1,23	.17	.01	.681

Table A.36: *Single Comparisons between bothNEAR and bothFAR Display Arrangements Using Paired t-tests*

Display	Task	bothNEAR		bothFAR						
		M	*SD*	*M*	*SD*	*d*	*t*	*df*	*p*	p_{corr}
Number	Different	5.42	5.79	3.27	3.97	.43	1.72	23	.050	.149
	Similar	2.54	4.33	0.75	2.10	.53	1.88	23	.037	.146
Pointer	Different	10.30	12.60	10.22	12.01	.01	0.04	23	.486	.486
	Similar	8.13	6.44	7.00	7.27	.16	0.68	23	.251	.503

Note. Error rates are in %.

Response Times

Table A.37. *Results of the Three-way GLM Repeated Measures Analysis*

Effect	*df*	*F*	Partial η^2	*p*
Display type***	1,23	326.78	.93	.000
Display position	1,23	4.23	.16	.051
Task***	1,23	67.32	.75	.000
Display type x Display position	1,23	.28	.01	.601
Display type x Task	1,23	.37	.02	.549
Display position x Task	1,23	1.14	.05	.297
Display type x Display position x Task	1,23	1.80	.07	.193

Table A.38. *Single Comparisons between bothNEAR and bothFAR Display Arrangements Using Paired t-tests*

Display	Task	bothNEAR *M*	bothNEAR *SD*	bothFAR *M*	bothFAR *SD*	*d*	*t*	*df*	*p*	p_{corr}
Number	Diff.	1205.98	221.10	1236.88	273.23	.12	0.74	23	.234	.468
	Sim.	958.65	216.12	1004.91	209.91	.22	1.48	23	.077	.230
Pointer	Diff.	2071.90	422.87	2214.14	214.86	.42	1.62	23	.059	.237
	Sim.	1856.90	354.65	1861.49	402.78	.01	0.08	23	.469	.469

Suitability Judgments

Table A.39. *Results of the Three-way GLM Repeated Measures Analysis*

Effect	*df*	*F*	Partial η^2	*p*
Display type***	1,23	25.68	.53	.000
Display position**	1,23	11.98	.34	.002
Task	1,23	1.10	.05	.304
Display type x Display position	1,23	.19	.01	.668
Display type x Task	1,23	.37	.02	.550
Display position x Task	1,23	.65	.03	.428
Display type x Display position x Task	1,23	.31	.01	.581

Table A.40. *Single Comparisons between bothNEAR and bothFAR Display Arrangements Using Paired t-tests*

Display	Task	bothNEAR *M*	bothNEAR *SD*	bothFAR *M*	bothFAR *SD*	*d*	*t*	*df*	*p*	p_{corr}
Number	Different*	0.80	0.61	0.39	0.68	.63	2.49	23	.010	.020
	Similar**	0.83	0.53	0.42	0.49	.80	3.37	23	.001	.005
Pointer	Different**	0.38	0.34	-0.04	0.59	.87	3.02	23	.003	.009
	Similar*	0.40	0.48	0.12	0.53	.55	1.89	23	.036	.036

A.5 Study 5: Reliability of FORCE image schemas

A.5.1 Definition of FORCE Image Schemas

Table A.41 contains the definitions, notations, and user interface examples that were given to participants in study 5. The graphic notations are based on Talmy's (1988, 2000) notational system of FORCE image schemas and visualisations by Johnson (1987) that were explained to the participants in the study.

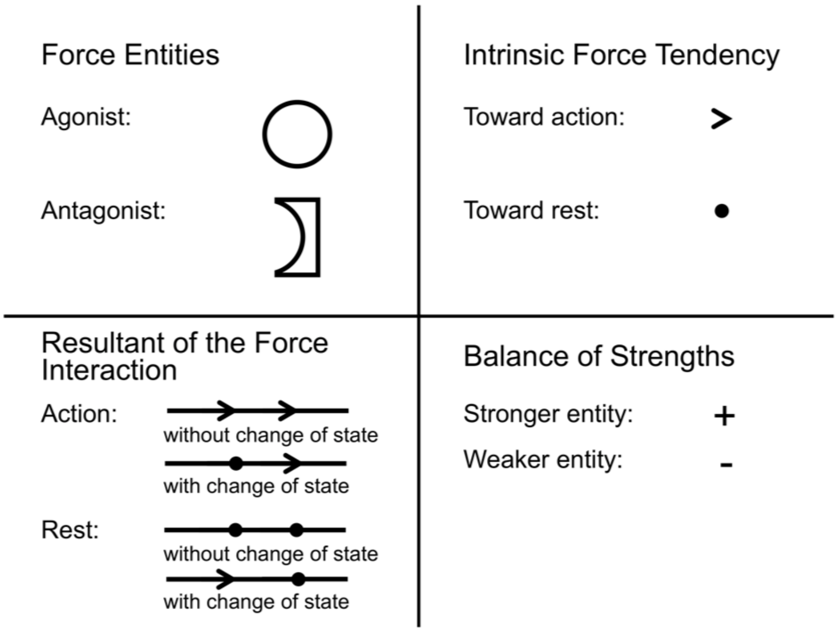


Figure A.1. Elements of Talmy's (1988) notation of FORCE dynamic elements.

Table A.41. *Definitions, Notations, and Examples of* FORCE *Image Schemas*

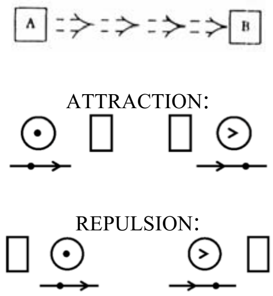
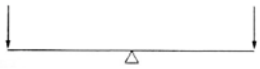
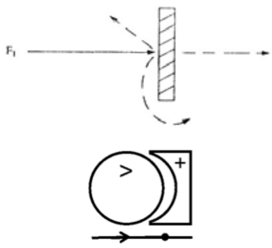
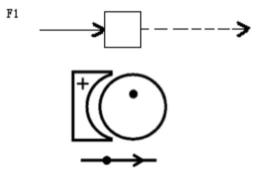
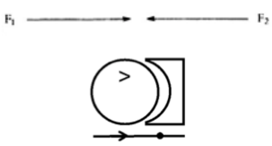
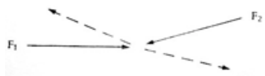
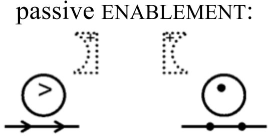
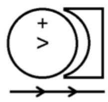
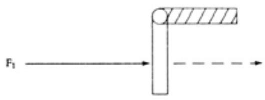
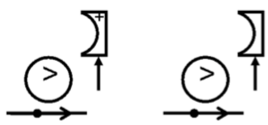
Image Schema	Definition	Notation (Johnson, 1987; Talmy, 1988)	User Interface Example
ATTRACTION / REPULSION	A FORCE image schema in which a (passive) object exerts a force on another object, either physically or metaphorically, to pull it toward itself (or in the case of REPULSION to repel it), mostly acting from a distance.		If the seatbelt is not fastened in the car, then a beeping sound is activated to alert the driver. (ATTRACTION)
BALANCE	A FORCE image schema that provides an understanding of physical or metaphorical counteracting forces: forces and/or weights counteract /balance off one another. Metaphorically, there is equilibrium, not *too much* and not *not enough*.		On both sides of the monitor screen of the cash machine there is the same number of equally sized push buttons. The symmetrical arrangement suggests similar functions of the buttons.
BLOCKAGE	A FORCE image schema in which a force / movement is physically or metaphorically stopped or redirected by an obstacle.		The car driver pulls on the handbrake to prevent inadvertent rolling.
COMPULSION	A FORCE image schema that involves an external force physically or metaphorically causing some passive entity to move.		The car driver steps on the accelerator and the car accelerates.
COUNTER-FORCE	A FORCE image schema that involves the active meeting of physically or metaphorically opposing forces that are equally strong. Both forces collide, there is no further movement.		The plane pilot wants to descend, the autopilot to ascend. Both struggle against each other. The plane neither descends nor ascends.

Table A.41. *Definitions,… (Continued)*

Image Schema	Definition	Notation (Johnson, 1987; Talmy, 1988)	User Interface Example
DIVERSION	A FORCE image schema that involves forces that physically or metaphorically meet and produce a change in direction or force vectors (at least one).		A user is checking her email and finds an interesting link. She follows the link and thereby loses sight of her actual work.
ENABLE-MENT	A FORCE image schema that involves having (a) the physical or metaphorical power to perform some act, or a potential force (vector) and the absence of BLOCKAGE, RESISTANCE, COUNTERFORCE, or COMPULSION; (b) a felt sense of power to perform some action	 passive ENABLEMENT:  active ENABLEMENT: 	When the car is taking a bend, the cornering light will actively light into the bend where the driver needs to look. (active ENABLEMENT)
MOMENTUM	A FORCE image schema that involves the tendency of an object to maintain the actual state of motion (or rest) if there is no influence of another agent.		The progress indicator of an mp3-player is moving as long as the song is playing or until it is stopped.
RESISTANCE	A FORCE image schema that involves a force that tends to oppose or retard the motion of another entity.		The shutter release button of a digital camera has a soft stop that triggers the auto-focus. When the user presses the button harder, the shutter is finally released.
RESTRAINT REMOVAL	A FORCE image schema that involves the physical or metaphorical removal of a barrier to the action of a force, or absence of a barrier that was potentially present.	 	The car driver releases the handbrake to move off.

In Talmy's (1988) system of "force-dynamics" there is always an Agonist and a stronger or weaker Antagonist. Agonists have either an intrinsic tendency toward rest or toward motion. Talmy developed a notational system that indicates the Agonist by a circle and the Antagonist by a concave form (*Figure A.1*). Further, the Agonist's intrinsic force tendency, the resultant of the force interaction, and whether an entity is stronger or weaker than the other is coded in the notation. An example of the notation of the BLOCKAGE image schema is given in *Figure A.2*. Here, the Agonist's tendency is towards action, but it is held back by a stronger Antagonist so that the Agonist is kept in place.

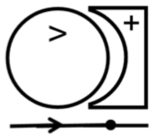


Figure A.2. Notation of the BLOCKAGE image schema, after Talmy (1988)

A.5.2 Pre-Study: Generation of Usage Examples

Table A.42. *Number of Generated Usage Scenarios and Confidence Scores*

Image Schema	Number of usage scenarios generated	Confidence Scores *M*	Confidence Scores *SD*
ATTRACTION/REPULSION	24	3.83	1.27
ENABLEMENT	18	4.13	1.02
BLOCKAGE	15	4.15	1.46
COMPULSION	14	4.36	1.28
DIVERSION	14	3.29	0.73
MOMENTUM	13	3.31	1.11
RESTRAINT REMOVAL	13	3.62	0.96
RESISTANCE	12	3.75	1.29
COUNTERFORCE	12	3.17	1.27
BALANCE	11	3.60	0.97
Total	146	3.74	1.19

Note. Confidence ratings ranged from 1 – very uncertain to 5 – very certain. The number of generated usage scenarios is higher for ATTRACTION/REPULSION because two questionnaire sheets were circulated for this image schema (one each for the other image schemas).

A.5.3 Main Study: Image-Schema Classifications

The classifications of each pair of raters were put into a matrix and added up over all possible pairs (Table A.43). The table contains the data of all six possible pairs out of four raters. Numbers in cells are absolute frequencies. Numbers in bold print denote agreement.

The classifications of each participant were also compared against the standard classification obtained in the pre-study and aggregated in a matrix (Table A.44). The table contains the combined data of four participants. Again, numbers are absolute frequencies and bold print denotes agreement.

Table A.43. *Observed Agreement and Disagreement of* FORCE *Image Schemas*

	AT	BA	BL	CF	CP	DI	EN	MO	RE	RR	Total
ATTRACTION	**63**										63
BALANCE	0	**31**									31
BLOCKAGE	0	0	**35**								35
COUNTERFORCE	3	2	5	**9**							19
COMPULSION	0	2	8	10	**25**						45
DIVERSION	2	0	6	1	1	**26**					36
ENABLEMENT	0	3	2	0	12	0	**47**				64
MOMENTUM	0	3	7	1	10	4	6	**22**			53
RESISTANCE	1	3	11	8	4	5	2	2	**12**		48
RESTR. REMOVAL	0	0	11	0	2	3	22	2	9	**25**	74
Total	69	44	85	29	54	38	77	26	21	25	468

Table A.44. *Observed Agreement and Disagreement between Single Raters and the Classifications Obtained from the Pre-Study*

Standard →	AT	BA	BL	CF	CP	DI	EN	MO	RE	RR	Total
ATTRACTION	**44**	0	0	0	0	0	0	0	0	0	44
BALANCE	0	**22**	0	1	0	0	0	1	1	0	25
BLOCKAGE	0	0	**27**	1	1	3	0	2	2	4	40
COUNTERFORCE	1	0	3	**6**	2	2	0	0	2	0	16
COMPULSION	0	1	4	2	**21**	3	0	1	2	0	34
DIVERSION	2	0	0	0	1	**20**	0	1	1	0	25
ENABLEMENT	0	1	0	0	5	0	**36**	1	1	5	49
MOMENTUM	0	0	1	1	1	2	2	**20**	0	0	27
RESISTANCE	1	0	1	1	0	1	0	1	**15**	3	23
RESTR. REMOVAL	0	0	0	0	4	1	4	0	0	**24**	33
Total	48	24	36	12	35	32	42	27	24	36	316

Table A.45. *Confidence Scores in the Main Study*

	Confidence Scores	
Image Schema	*M*	*SD*
ATTRACTION/REPULSION	4.33	0.41
RESTRAINT REMOVAL	4.22	0.45
COMPULSION	3.83	0.92
BALANCE	3.47	0.98
DIVERSION	3.46	1.27
RESISTANCE	3.39	0.79
BLOCKAGE	3.36	1.18
ENABLEMENT	2.94	1.06
COUNTERFORCE	2.58	1.25
MOMENTUM	2.50	1.23
Total	3.40	1.12

Note. Confidence ratings ranged from 1 – very uncertain to 5 – very certain.

A.6 Study 7: Redesign of a Business Application

Table A.46. *Comparisons of the Results of the AttrakDiff Questionnaire Using the Fisher-Pitman Randomisation Test*

	SAP R/3		IS-GUI		IS-HYB				
Scale	*M*	*SD*	*M*	*SD*	*M*	*SD*	$p_{SAP\text{-}GUI}$	$p_{SAP\text{-}HYB}$	$p_{GUI\text{-}HYB}$
Pragmatic	0.29	0.56	1.74	0.57	1.23	1.00	.016	.071	.401
Hedonic-Stim.	0.00	0.39	0.89	0.88	2.17	0.62	.063	.008	.016
Hedonic-Ident.	0.25	0.49	0.86	0.77	0.94	1.30	.111	.222	.469
Attractiveness	0.18	0.49	1.26	1.01	1.29	0.88	.056	.024	.484

Note. SAP R/3 is the current system ($N = 4$), IS-GUI is the graphical user interface, IS-HYB is the hybrid graphical-tangible user interface ($N = 5$).