CHAPTER **U**

User Behavior

On Friday morning, Henri's alarm clock rings at 7:00. His firm intention is to make some real progress with his assignment work. But before starting, he takes a long walk and tries to clarify his ideas. Someone has told him that during the nineteenth century all students in a famous university had to do some physical exercise two hours in the middle of each day. Henri has also noticed that during apt physical exercise he was more creative and less fixed to a regular way of thinking. Henri walks almost an hour although the weather was pretty freezing. He does not put much attention to external conditions, but rather ruminates on the most urgent matter for him, that is, Irene. After a decent sleep, he is ready to accept and even appreciate the fact that ladies often are sensitive and spontaneous. Now Henri has no doubt about his willingness to put a considerable effort to improve their relationship. But what could he, an ordinary engineering student, do?

Henri's thoughts go back to the topic of the lecture on Wednesday. Dr. Leopard's advice for men is to give roses to their ladies. It would be easy to be cynical, he had thought, and judge those kinds of small deeds as naïve, calculating, or selfish. Maybe so, but Dr. Leopard asserts that proper small deeds could be used to build a flourishing long-term relationship. Perhaps, Henri thinks, he also should strengthen his own bond with Irene by doing something similar. A bunch of roses ordered for Irene in Milano—that would be nice. On their early days of relationship, he had once, without any serious plan, given a bunch of flowers to Irene, and the effect had been marvelous.

Nevertheless, Henri is an engineer by heart. While walking towards his apartment, he starts to think whether something similar could be achieved by means of IT technology. *What does make flowers so stimulating? They are genuine, yes, and lovely, and perhaps even somehow energetic.* Henri thinks—properly for a student with high grades in elementary physics—that it is somewhat strange to think that flowers can yield energy. Either energy is in a measurable form, or there is no energy at all. Still, by recalling the occasion when Irene got the flowers, he can feel something that could be called energy, though not necessarily measurable by the devices used in the laboratory of physics.

Henri begins to get inspired. A feeling of energy, or whatever it should be called, is something real, in a way more real than anything else because it can be perceived directly while any other thing is a just a result of interpretation. *If I am really angry, I can hardly have any doubt about my angriness, whereas I can always question* *whether the tree I just now stare at is real or not. The same is true with delight. The observed prospect of improving my own or my friend's mood by a flower or by an IT application creates a need to do something. For instance, I might be willing to buy a flower or a new application because of the expected benefit.* This prospect forms a kind of driving force that may even provide the energy needed to sustain a chain of actions, possibly even an ecosystem. Irene often used the term ecosystem in various situations. Henri had just a couple of days ago pondered what a design ecosystem really means, as for him ecosystem appears to refer to biology, animals, plants, and so on.

Surely, I will not aim at becoming a florist, Henri thinks. Instead, I may try to invent something more technical that is able to create delight. Or innovate—anyway, what is the difference between invention and innovation? An invention can be transformed to an innovation, or was it vice versa? Henri has heard something about the difference during a business-related course, but had not paid much attention to the issue.

Henri returns to the question about real implementation of his invention. Can it be just a piece of software or does it need some special hardware and what should be the logic used in the application? He has, similarly to most people that have used smart phones, struggled at the beginning to understand what the developers of the phone have really thought. Sometimes it has been notoriously difficult to use new phone or a new application. Henri wants, as a minimum, to avoid the worst mistakes of new gadgets and widgets.

He became interested in the design principles and processes mainly thanks to Irene's enthusiasm on style and design. Now he even starts to believe that anything that is designed shall be designed not only well but also remarkably well. Maybe there was not so much difference between a well-designed phone and an artistic product or fashion, or even something created through evolution.

Irene had also tried to explain the concept of conceptual model. It was somewhat hard for Henri to distinguish the overt logic of a device and the conceptual model each user has in his mind about the logic the device. Still, after a systematic effort he now keeps those two concepts separate in his mind. Now he thinks that the user must always rely on his own conceptual model when he uses a device. The point is that the device shall offer understandable clues to how it shall be used. Besides, also Henri uses manuals only as a last resort.

So, Henri notices that he needs an idea and a name. Something that combines emotional and technical aspects, maybe an application that guides users from negativity to positivity – Flourishator appears in his mind when he arrives at his apartment. He quickly checks whether anyone has used that term. No relevant hits so far, great! His cunning idea is that Flourishator may attract also male users that are wary with anything that creates a too emotional impression. He also notices that the starting figure of the application must be carefully designed in a way that it is acceptable both for female and male users.

Henri spends the whole evening to write a draft of the design of Flourishator. He has some rudimentary idea of a tool that makes solid proposals to improve the well-being of the user. A possible starting point is a pair of emotions: the first one is the current emotions felt by the user, and the second one is the emotion towards which the user wants to proceed. He imagines some fifteen names of emotions arranged in the form of a wheel: anger, happy, sad, excited, etc. The user first moves his finger on the display to select the current state of his mind. Then the user selects a desired state of mind by moving his finger on top of another emotion. Finally, the application gives a proposal what to do to make the change happen. Simple as anything. But there are still some open questions, Henri ponders.

Who is able to give a credible proposal for cases in which, say, an angry person wants to become happy? The only idea that comes in Henri's mind is that Flourishator shall recommend the user to play Angry Birds. Yet there have to be hundreds of somehow beneficial proposals to make the application permanently useful. It does not make any sense to give the same advice every time the user wants to be happy, because then the user will become bored or frustrated rather than happy.

Henri is actually eager to consult Irene and her colleagues about the design of Flourishator. The use of the application has to be very intuitive, without any user manual or additional instructions whatsoever. Signifier—that was the term Irene has regularly used. Furthermore, user experience has to be smooth and pleasant, with some nice visual tricks to make the product attractive from the start. And no error messages, ever.

Just now, his phone alerts him with the familiar tone that was reserved for Irene. He glimpses the clock on his laptop screen, 16:12. He was already expecting the call. He closes the laptop and picks up the phone.

Content

User behavior is one of those topics that we are familiar with but without any systematic insight on the topic. We use technical devices but are not aware of the typical characteristics of our own behavior. Most of the time, we do not pay any attention to the issue of how we are using devices, and when we do pay attention, we are typically dissatisfied or frustrated with some properties of the devices. When we are frustrated, we concentrate on the specific situation and try to solve the problem at hand.

This chapter consists of five main sections. First, activity theory is introduced briefly. The main strength of activity theory is that it provides a simple framework and an efficient terminology to observe and model the behavior of everyday life. Then a usability example is presented. The main point of the example is to highlight the connection between a typical example of using a novel device and the theoretical parts of this book. Secondly, usability is addressed as an important topic that may even define the success of a communications ecosystem. Thirdly, the issue of measuring the perceived quality of service is addressed. The section underlines the fact that the terms used to assess quality should be selected very carefully. Fourthly, a web-browsing type of application is discussed and modeled in a situation where the available bit rate is changing and thus affecting the benefit obtained by the users. Finally, some models are introduced to consider the network effect, that is, how the number of users or service penetration affects the benefit obtained by a user of the service.

Terms

As user behavior is an intermediate area between psychology and technology, most of the key terms provide a linkage between human life and technology: needs and motives are the drivers of human activities while applications and the Internet provide technical means to satisfy needs and motives. Usability is one of the key terms that makes a connection between technology and users.

As a communications ecosystem expert (CEE) you shall remember the fundamental difference between World Wide Web (WWW, or briefly web) and the Internet. The Web is a system consisting of interlinked documents whereas the Internet is a system consisting of physical networks, links, and nodes. Although they often are mixed in informal text or speech, a CEE shall be able to distinguish them properly. Other terms include:

action: an event in which something is done so as to accomplish a purpose,

activity: a purposeful interaction of a subject with the world,

agent: an entity that is able to strive for its own goals,

application: a computer program that provides a user with tools to accomplish a task,

behavior: the actions or reactions of a person in response to external or internal stimuli,

concept: an abstract idea describing a piece of reality,

context: the circumstances that form the environment within which something exists or takes place,

design: realization of a concept or idea into a configuration, drawing, model, pattern, plan or specification,

motive: a drive, force, or tension state within the organism that impels it to act,

need: a physiological or psychological requirement for the well-being of an organism,

role: the characteristic behavior pattern of a person in a particular context, social setting or environment,

usability: an attribute of a product or service that describes how easily users can perform tasks required to achieve the expected benefits,

user: a person who makes use of a thing, and

world wide web: a system of interlinked hypertext documents accessed via the Internet.

Other terms in this area of user behavior include:

ability	decision	MOS	signifier
access rate	effort	multimodal interaction	social media
activity theory	event	network effect	stimulus
actor	flow	object	success
aesthetic	fulfillment	operation	task
attempt	functionality	opportunity	usage
competence	HCI	peer to peer	usefulness
conceptual model	incentive	reaction	user experience
convenience	irrational	safety	user satisfaction
creeping featurism	message	session	VoIP

Activity theory

It would be helpful for a CEE to be familiar with activity theory. It provides a systematic framework "to understand the unity of consciousness and activity" as Kaptelinin and Nardi (2006, p. 8) express the objective of activity theory. In the framework of this book, *consciousness* might be replaced by the term *mind*. Mind, which contains many subconscious processes, creates motives while an activity is the instrument to accomplish the objectives determined by mind. Someone may even express the same idea as follows: after a motive has emerged in mind, an activity is selected either consciously or unconsciously. It may even be that the "real" motive for an activity does not necessarily reach consciousness at all. In contrast, conscious thinking might be used to invent or construct a socially acceptable reason for the actions related to an activity with a hidden, non-conscious motive. Consciousness is not used so much for defining what we genuinely want but for defining what we want others to believe about ourselves.

Activity, action, operation, motive, goal, and condition form the core terminology for activity theory as illustrated in Figure U.1. The terms are used similarly in this book. I will use also the term *session* to describe a sequence of operations made by a user that appears to form a consistent and continuous event from a technical perspective. Typically, a (technical) session corresponds to (human) action, but in some cases, even an operation can induce a session viewed from technical perspective. A session might also consist of a series of actions. In the technical or service domain, the total usage of an application (like text message) corresponds to activity in activity theory. However, we cannot assume any one-to-one correspondence between an activity and the usage of an application, because each application can be used to support various activities and various motives, and many applications can be used during an activity.

A similar hierarchical structure is also used in the domain of Experience design (see Hassenzahl 2010, p. 44). In experience design "be goals" correspond to motives, "do goals" correspond to goals, and "motor goals" correspond to conditions. Needs and missions are located above all of these constructions. The total benefit of using communications services and applications can be thought to be on the level of missions and needs.

We may also consider the structures presented in Figure U.1 from the viewpoint of Abraham Maslow's Theory of Human Motivation (1943). Maslow has been criticized as concentrating too much on individuals and their needs at the expense of society. However, even Maslow makes the following remark:

Motivation theory is not synonymous with behavior theory. The motivations are only one class of determinants of behavior. While behavior is almost always motivated, it is also almost always biologically, culturally and situationally determined as well.

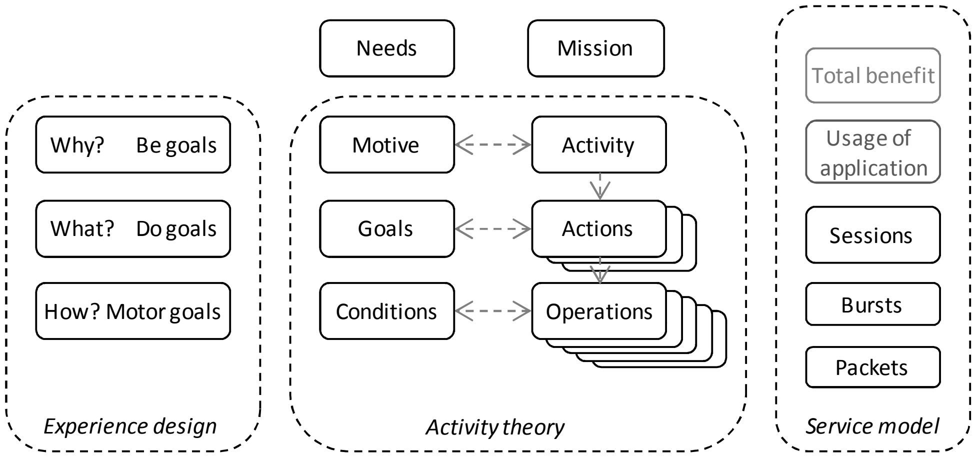

Figure U.1: The hierarchical structure of activity according to experience design (Hassenzahl 2010), activity theory (Kaptelinin and Nardi 2006), and the communications service model used in this book.

Although the mathematical models describing the usage of services are almost inevitably built on the assumption that each user somehow fulfills his or her personal needs we need to be aware of the fundamental limitations of those kinds of models. We do not make decisions only based on our own, clear defined and observable needs; instead, our behavior is affected by numerous external factors often in a way that we are not able to detect.

The main points to be stressed here is that when users make decisions about the usage of communications services the decisions are made in the context of an activity, and the decisions are not always based on any conscious reasoning.

When we model the behavior of users sending text messages, we need to remember that the user assesses the benefit of the text message in the light of an activity. A typical motive for an activity is "to maintain a good relationship with another person." An action of the activity would be to initiate a brief conversation with the other person by using text messages. From technical viewpoint each text message forms a session, while from user viewpoint an action consists of a set of sent and received messages. Particularly, when a person sends a brief message using special language and abbreviations with highly automated maneuvers, the message can be situated on the level of operations. In those cases, a text message has a similar value as one sentence in a book. There is a flower that bloometh. An unfit message or sentence may interrupt the effortless flow of an action or reading and engender considerable frustration, while a fitting message or sentence does not cause any noticeable emotion before the goal of the action has been fulfilled. This has a significant implication on the way we are thinking about the benefit of a session. The real benefit is typically achieved on the level of activity. A session (for instance, sending a text message) serves the purpose of the activity but is often meaningless when assessed out of context. If a session disturbs an action, or in worst case, jeopardizes the whole activity, the cost can be many times larger than the benefit of a successful session

The objective of the bloometh sentence in the middle of previous paragraph (if you paid any attention to it) was to give a concrete example of an unfit sentence (the sentence is from Ulysses by James Joyce, which means that bloometh may refer to Mr. Bloom, the main character of the book). What might have happened is that you continued reading until a subconscious process interrupted the ordinary reading process and started a higher-level process to consider the reasonability of the sentence. That kind of interruption induces a mental cost, but I hope that the cost in this case was acceptable in light of understanding the user behavior and emotions.

In general, any interruption during an action is painful, if the user needs to take a step back and look the situation from a metalevel perspective in order to understand what is going on, instead of concentrating on the action itself (see Kaptelinin and Nardi 2006, p. 259). This phenomenon also explains why people are not willing to invest in a solution that might in the long term save time and effort, if the investment disturbs the progress of the current action. The prospective benefit does not exceed the cost of present annoyance, because you cannot make an objective analysis about the pros and cons without interrupting the ongoing progress of the action.

Thus, there are different requirements for continuance of events on different levels: we can handle numerous activities at the same time, whereas we are poor in performing several unrelated operations at the same time. Automatization may yet improve our multi-tasking skills. For instance, an experienced driver might be able to drive a car and converse with a friend at the same time, if the driving environment is simple and clear. Still, as simple task as sending a text message is a risky effort when driving a car. Operations shall be done in succession and without any interruption. As to actions, although interruptions are possible, they should be avoided whenever possible because they consume extra resources. Consider, for instance, a simple scheme in which you have a free phone service except if the network becomes congested. Then you had to decide whether you are willing to pay a small price to continue the call. The mental effort or cognitive load of considering an unrelated question during a phone call is probably too high to make the scheme feasible.

Another aspect of actions is that if we attempt to understand the reason for an action of a person (to be able to assess the benefits it provides) we are forced to take the viewpoint of an outside observer instead the viewpoint of the real decision maker. There are two options to solve this dilemma. The first option is to remain as an objective, neutral observer and to use observations, measurements, and formal reasoning to assess the situation. However, that option tends to give too much emphasis on objectively measurable things (like money) compared to the inner motives of the person. In the second option, we have to give up the role of neutral observer and instead situate ourselves in the place of the actor and utilize the feelings that the imagined situation produces.

Even though in most cases I would prefer the second option, we need to be aware of our tendency to explain our emotions and behavior in a socially acceptable manner. It is hard for me to imagine a situation in which I would do something socially unacceptable, like disseminate computer viruses. Activity theory may help here by pushing us to understand the motives of the activity (for instance, producing computer viruses). If the motives are understood well enough, the goals of separate actions and the methods of operations will be more comprehensible. Thus, the activity theory claims that any reasonable analysis of user behavior shall take into account the context in which the user interacts with the world. A laboratory environment is a specific context for the users, which means that the results of an experiment do not necessarily reflect the behavior of the same users in more realistic environment.

Another argument of the activity theory is that we need to take into account the asymmetry between people and things. Activity theory states that only living things have needs and only humans have intentions (whether some animals, e.g., dolphins express intentional behavior, is out of the scope of this book).

Thus although technical devices and systems can be active and maybe even somehow intelligent (stupid they can surely be), they do not have their own needs or intentions. It may appear that the intention of an autopilot is to keep an airplane stable and directed towards a pre-defined target. The system may even be considered intelligent if it is able to handle many problematic situations, for instance, quickly changing winds. However, this kind of expression would represent a misuse of the concept of *intention*. The intention of the designer of the autopilot might have been to design a system that is able to behave similar to a real pilot with certain intentions. The autopilot still does what its realization forces it to do, without any intention of its own.

Similarly, we should avoid of using the terms cognition and cognitive in the context of artificial systems (see also Kaptelinin and Nardi 2006, p. 204). Instead, it would be more consistent to use term *functional properties* in case of technical systems. Of course, it is easy to admit the attractiveness of terms like *Intelligent Robot* or *Cognitive Radio*. One may speculate that those terms have originally selected because of the need to get funding for research and development projects. They may have served well that purpose. Still the terms are problematic because they suggest that the technical systems behave similarly as human beings, with similar abilities, needs, and intentions. Still they are not human beings, instead they are technical artifacts, perhaps complex and hard to understand, but often also very useful for the purpose they are designed. I hope that you will make a clear distinction between human beings and artificial things also during a career of CEE. As to non-human animals, opinions vary whether or not they should be assessed and treated similarly to humans.

As a summary, we may make the following statements:

- Everything, even non-living things, can have an effect on other things.
- All animals can *act* according to *biological* needs, and
  - maybe also plants.
- Human beings and social systems can act according to *cultural* needs, and
  - maybe also some animals living in large groups with extensive social interaction, like chimpanzees, elephants, and dolphins.
- Human beings and social systems can perform *intentional* actions,
  - still, some recent findings indicate that some animals act as if they have intentions.
- In addition to humans and social systems, also devices and cultural animals (e.g., horses and dogs) can be used to realize the intentions of human beings.

Another type of summary is presented in Table U.1. It shows how different parts of a system or an ecosystem are used and assessed at different phases of development or evolution. The table is partly based on Kaptelinin and Nardi (2006, p. 101), but the metrics columns are added.

Furthermore, we can notice that *expansive* can mean either intentional development of something or an evolutionary process in which various sub-processes pick up and discard different choices. Thus, it is not at all clear that the result will be good even with regard to the selected metrics. Particularly organizations and communities tend to evolve independently of the intentionally made plans. Different types of metrics adopted by different actors may result in direct conflicts between the actors. For instance, technical people may use performance as a metric whereas socially oriented people may use worth as a metric.

The main point I would like to stress is that the metrics used when we assess an entity that is already in permanent use typically differs from the metrics used to assess any further development of the same entity. Think, for instance, rules that are used to control our behavior when driving car, like the speed limit. When thinking of a rule that has been in place for a long time, we typically consider how well drivers are performing in everyday situations when they obey (or do not obey) the rule. In that case, we often avoid making questions about the wider effects of the rule. For instance, when everyone drives approximately the same speed, it is relatively easy for everyone to predict how the state of traffic will evolve in the near future. Even one driver with much higher (or lower) speed can significantly impair that predictability and thus seriously disturb the fluency of traffic and safety of other drivers. Even if the speeding driver is exceptionally skilled and can avoid accidents, he may have negative effect on other people. Still he may consider himself entitled for speeding because of his superior skills.

In contrast, those who develop the rules should consider whether the rule is worthy of implementing from the viewpoint of the society, taking into account both all relevant cost factors (including people's resistance against strict rules and monetary cost of control) and all beneficial factors, particularly saving human lives.

Table U.1: Parts of system supporting an activity presented from the viewpoint of metrics.

	Phase to be assessed			*Typical metric for*	
	Predetermined	*Active*	*Expansive*	*Past and fixed*	*Future and open*
Tools	Automated usage	Conscious usage of a tool	Tool development	Capability	Efficiency
Organization of work	Fixed division of responsibilities	Mutual coordination	Development of organization	Performance	Efficiency
Rules to control behavior	Fixed control	Shared meanings and objectives	Rule construction	Performance	Worth
Actions: making decisions	Triggering automated action	Processing information to make better decisions	Learning, gaining insight	Benefit	Worth
Organization of community	Fixed hierarchy	Flexible network	Community construction	Efficiency	Worth

Similarly, we assess a tool based on the capabilities it offers. However, if someone develops a tool for himself he probably considers primarily the benefits it offers for himself compared to the cost it produces. That ought to be case also when developing commercial products. One may argue that it would be too mentally burdensome to assess in case of an individual action whether the action is really worth doing. Thus even though *worth* is somehow more important and more fundamental metric than *capability* or *performance*, it is often reasonable to limit the assessment of the consequences of an action on these lower levels, when the consequences likely are minor. It is more reasonable to make a *worth* assessment on the level of activities.

In Milano, Irene wakes up somewhat late on Friday morning. She is not as brisk as usual, because she had slept poorly after a trifling quarrel with Henri. Henri's attitude had really annoyed Irene. She had said, using as soft and feminine tone as possible that she was feeling lonely. And Henri didn't react at all, he had just continued his story about the lecture of Dr. Leopard without giving any notice. Her tone had worked pretty well in the early days of their relationship. Besides, Henri had not even remembered Women's day. She had waited the whole day. Fortunately, some of his male friends at the Design Institute were much more well-mannered. Even the last text message from Henri just before midnight was a mere disappointment. Maybe Henri has indeed devised an automatic text message generator as he had joked some weeks ago.

Irene starts to get irritated when musing over all the incidents. Perhaps they were symptoms that were trying to tell her something important. Anyhow, Henri surely did not really appreciate her feelings. What had changed? One year ago, she had been more excited than ever.

On a nice spring afternoon, Henri and she had taken cappuccinos in the small cafeteria on the ground floor of University's main building. For some reason, the cafeteria was much more crowded than usual. There was hardly any empty seat available, but she had felt an energetic and encouraging atmosphere. Irene had wondered whether something extraordinary was happening there. Then Irene had noticed a small group of older people—at least compared to ordinary students—discussing about something interesting. Regardless of the noise in the cafeteria, she had discerned miscellaneous words, such as design, magnificent, and Italy. Oh, Italy, what a wonderful country, she had always dreamed of living there.

So, she had asked politely whether they could sit at their table, while Henri had looked precautious. As she liked to use Italian phrases she began with gently articulated "Buona sera." Irene was eager to know what all these people were doing in the cafeteria and why they were so excited. The youngest member in the group, a handsome gentleman told that they were on a break of lecture. She had also noticed the shining shoes and carefully selected necktie—he was almost like an Italian. Irene still remembers vividly how Henri had looked suspicious at that moment. It was the last lecture of a course, continued the lady, and it was not about rocket science but a series of lectures on philosophy and systems intelligence. This lecture was about dreams. Amazing, Irene thought, and it seemed that even Henri began to pay some attention to the discussion although he remained silent.

"You had an interesting discussion going on before we interrupted your discussion," Irene said in a way that revealed her curiosity. The lady told that she and the older gentleman worked at a small design company while the younger gentleman was responsible for developing cooperative projects at Aalto. They were discussing about the cooperation that has been taken place between the company and a design institute in Milano. Their common dream was to establish a new curriculum that is able to integrate the essence of several disciplines and to combine the strengths of northern and southern Europe. They were just starting a student exchange program between Aalto and Milano. This small episode later resulted in Irene's visit to Italy.

Besides, although they had some other plans Irene was able to persuade Henri to come with her to listen to Dr. Leopard's last remarks. These were cherished memories for Irene.

She had loved the joyful and weightless atmosphere in Milano, but at the same time, she yearned for the more down-to-earth attitude of Finnish culture. Thus she was pretty annoyed when Henri had yesterday mentioned the cafeteria instance, and said that maybe they should not visit too often in the cafeteria to avoid all kinds of new adventures. What did Henri mean by that? Should she sacrifice her dreams and stay always in Finland even during the cold winter months—she likes sun and warmth. She is sure that Henri will never really understand her aspirations and dreams. She starts to feel lonesome and depressed. Better to forget Henri, she resolves and makes a call to an Italian friend. She becomes a little bit revived and suggests that they should go to the nearby shopping center. A new blouse or a glass of wine would refresh her more, and talking with a friend may help her to ignore her gloomy thoughts.

But just when Irene is leaving her apartment someone is on her door—she is not waiting for anyone, strange, she looks her stylish wristwatch; it was 4:00 pm. She hesitates for a while before opening the door.

What? A deliveryman hands over a big bunch of roses with a card. She was very stunned. She reads the card: "Ti amo per sempre."

Had Henri really had the same concern as I—miraculous incident, does he read my thoughts? There are no accidents, she tends to think, and there always is a pattern or a message, visible or invisible.

There was something else on the card as well. They had met at the midnight before the first May at a noisy student happening. Now, her darling tells that they have spent 16 384 hours together and that means 100000000000000 hours in binary code. A rose for every zero, 14 all together, and she is the number one, anyway. She calls Henri right away.

Usability

Technical devices extend our physical capabilities, often as extensions to our hand. A hand is something that we, naturally, experience as an integral part of our body. A healthy person does not usually complain about the poor usability of his own hand. We know, at least in principle, how to use a hand: all the processes between a conscious decision to bend a finger and the contraction of correct muscles take place automatically. Actually it is quite an astonishing thing to observe how own finger bends just because of a mental decision. What a perfect usability! In contrast, the use of digital devices typically requires specific instructions and a learning phase with conscious thinking. Therefore, there also is a comprehensive discipline called Human Computer Interaction (HCI).

What is the main difference between human computer interaction and mind-finger interaction? First of all, I possess my own fingers and no one else can control them directly while any device owned by me can be used by other people. Secondly, there is no significant change in user interfaces or operating systems: fingers can be controlled in the same way tens of years which definitely not the case with modern technology. Fingers are always available without searching for them, unlike many small devices. We do not need to charge batteries to keep fingers in operation.

Every device or tool requires an interface for controlling purposes. In some cases, the interface is so simple that we do not need to think of it. For instance, the use of a hammer is obvious. Still the activity during which the hammer is used can be very demanding. Norman (2011, p. 41 - 46) presents an illuminating example of silversmith's tools. Although the use of a silversmith's hammer is in principle simple, the selection of a fitting tool for each operation is a challenge, and, of course, a desired outcome requires exceptional skills even with the best available tool.

A hammer is a kind of analogical tool without any on/off buttons or displays. Many of our contemporary tools used for simple tasks include digital controllers. Figure U.2 shows the push buttons of the toaster I have used a couple of years. The toaster is quite easy to use and does not require significant mental effort during ordinary use. Still the meanings of the buttons are not entirely apparent. I glanced at the manual after I bought the toaster, but still I use only two of the four buttons. I made also a small survey among 29 students. The students were asked to define or guess the purpose of each button. They were told that the buttons belong to a toaster with a carriage lever, an adjustable slot lever, and a time controller. The purpose of the buttons and the survey results are explained later in this section in order to enable you to consider the same question about the meaning of the signs.

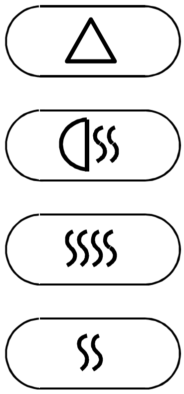

Figure U.2: Four buttons to control a toaster.

The purpose of Figure U.2 is not to point out any problems of this specific user interface, particularly because the array of buttons provides aesthetic value. Instead, the main objective is to illuminate the challenges of user interface design even when the purpose of a device is as simple as toasting slices of bread. Perhaps also a mobile phone could be designed by only using four buttons. Four buttons, if they can be pressed at the same time, enable $2^4 = 16$ different combinations. That surely provides enough capabilities to construct as complex system as wanted. Yet the usability might become awful because the automatization of actions would be an extremely painstaking process if the interface is unable to provide a comprehensible sign or

signifier: a signal in a physical or social world that can be interpreted meaningfully.

With complex operating systems and touch screens, it is possible to add huge amount of features without adding any control button. More often than not, this possibility results in

> **creeping featurism:** a tendency for systems to become more complex over time as more features are added than were in the original design or plan.

To take a handy example, you could even blame this book on creeping featurism (or featuritis) because of the large variety of topics, concepts, models, frameworks, formulas, examples, recommendations, definitions, and turns in the story. Pure text telling facts would appear much simpler. It is even said that each additional formula in a book halves the number of readers. However, we can consider the issue of complexity and features from other viewpoints. First, there are matters that are inherently complex and ecosystems definitely belong to that category. As an ecosystem expert, you need a variety of tools, as silversmiths do, too. It would be nice to own one CEE tool with a simple user interface suitable for all imaginable purposes, but that is not a realistic idea. Moreover, the mastering of a large variety of proper tools is the special asset of an expert, because there is a real need for expertise only when matters are complicated and challenging.

Similarly, our communication needs are so diverse and delicate that it is not likely that one tool could ever fulfill all needs. Anyway, human beings use many communication channels, which also means that multi-purpose tools must support multimodal interaction. Multimodal interaction, satisfaction of various needs and ever-increasing customer expectations may lead to a situation in which services become awfully complex and hard to design. Then the crucial question is: how could we design complex services in a way that they serve the main needs of customers? The answer given by Donald A. Norman (2011, p. 148) is:

> "The only way to solve the complexities of services is to treat them as systems, to design the entire experience as a whole."

Here we should also remember the definition of

> **service:** an event in which an entity takes the responsibility that something desirable happens on the behalf of another entity.

An excellent service is a service that takes the full responsibility of a complex matter. This cannot be achieved if the designer concentrates on separate symptoms without understanding the big picture. Norman (2011) uses train service as an example. If customers complain about the experience when travelling by train, it does not mean that the real problem can be solved by redesigning the interior of the train. The design of the service must cover the whole experience starting from the learning phase about routes and timetables, and ending with the events that occur after the journey. Similarly, if users are complaining about the experience of communications services, it is not enough to redesign the visible interface, although that most probably will be a part of the solution.

The results of the student survey were: 15 out of 29 students guessed correctly the meaning of the buttons (from top): cancel, one-sided heating (called bagel button by the manufacturer), high power, and low power. So far, I have used only cancel and high power buttons, but from now on, I may use all buttons.

I could not remember correctly the purpose of the two other buttons before reading the manual. Bagels I do not toast too often, but why did I forget the low power option? An explanation is that I wished to use the time controller as automatically as possible. An additional level of power (low) would at least double the difficulties of the automatization and make bad mistakes much more probable. We encounter the same problem with microwave ovens. A seemingly useful feature, like adjusting power, might also create negative side effects. Note also that initial sales might be maximized by a large number of features while repeat sales might be maximized by a considerably smaller number of features as argued and modeled by Rust et al. (2006).

A use case about usability

This section offers an account of some adoption and usability issues based on my own experiences. The main objective is to demonstrate that personal experiences can and should be used to understand more general phenomena, for instance, the process of being acquainted with a new device. Thus, my recommendation for those that decide to read this section is to consider their own experience in similar situations in parallel with reading.

In the middle of writing this chapter, I bought a new mobile phone. Most of the applications in the new phone were similar to those in my previous phone, except the GPS navigator. I already had used a phone for web surfing for several years, although only to a limited extent due to the small display and limited access rate. There were a couple of significant changes in the new device as well. The shift from a small keyboard to a touch screen made the change somewhat arduous during the first hours of usage. Furthermore, the vendor had made some significant changes in managing and controlling the phone and its applications. Finally, some new terms used in the user interface were difficult to comprehend and generated a considerable cognitive load during the learning process.

I had used a touch screen in several occasions including mobile phones and other devices. Still it took some hours to reach a level of expertise on which I felt confident with the new user interface. As discussed in Chapter H, the automatization of recurring operations is crucial for efficient completion of actions because of the slow and effortful nature of conscious thinking. An acceptable level of convenience can only be achieved through persistent usage of a new device, and that process requires sufficient motivation. I surely was motivated enough to go through the learning process as far as the touch screen control was concerned. Touch screen handling was easy enough to generate some positive emotions from the start including some pleasure due to the improved feeling of control.

A touch screen lead naturally to a different control structure compared to older phones with traditional user interfaces. However, I could still recognize some features that seemed to stem from the older phones without a touch screen. The complex mixture of features and principles made it hard for me to build a consistent conceptual model that could have guided through new territories inside the phone. As a result, even after a couple of months of usage there are numerous features and properties that I have never used. Separate *actions* of type "now I check how to download music to the phone" do not to promise high enough short-term benefits to enforce me to make any significant changes in my usage behavior. The only way to improve my skills seems to be find high enough motive to commence an *activity* during which I would systematically study the capabilities of the new phone.

I almost never use any automated synchronization. The reason for this habit, in addition to some early unpleasant experiences, is my wish to be in full control of the content of my personal data. As a result, I spent 31 minutes to move important events from the old phone to the new phone, manually. Whether or not that was an efficient action, is a debatable issue, but at least I avoided any frustration caused by the possible troubles of automatic synchronization. Besides, there was more than just the moving of the events: I was able to think about each event consciously and I even added some new information to some events. Note also that thinking about certain future events may induce happiness. An automatic calendar synchronization might well have generated frustration, and in the best case, satisfaction when the task had been completed.

Phone calls with the new phone did not cause any perceivable problem. According to my first experience, the voice quality is somewhat better than in my previous phone. Nevertheless, it is difficult to assess whether this experience was due to real improvement in audio quality or was it caused by the overall experience with the new phone. Still, the *experience* of improvement was real. As to text messages I did not expect any difficulties. Still, I had some problems finding the folder of received messages. The trouble generated a nuisance lasting roughly ten minutes.

My highest expectations for the new phone concerned the web browsing experience and navigation. Although I had used navigators occasionally, I was not an experienced user. The experience of getting familiar with the navigation tool consisted of various emotions. Some operations and even actions went smoothly and without any considerable stress. Yet there was an operation that caused a brief but quite intensive frustration: I was not able to immediately discover the sequence of operations needed to create a new favorite place. The learning process took about ten minutes (frustration naturally slows down rather than accelerates the learning process). After a couple of times of usage, I knew how to manage that operation but still I had to think of the operation. Frustration was clearly the most important negative emotion during the learning process. There was even a hint of shame because was not able to control the phone in spite of some expertise in this area.

As to the positive side, there were several emotions. First, I felt light pleasure when the actions run smoothly, and even excitement when the navigator demonstrated to be better designed than what I had supposed. Finally, I felt satisfaction when I finally was able to reach my targets. The positive emotions barely compensated the effect of frustration during the early days of usage. If you just add the arrows shown in Figure U.3 describing my emotions related to the first sessions with the navigator, the result is a small arrow to the right. However, because negative emotions typically have stronger effect than positive emotions, it might be more correct to say that the emotions were dominated by frustration. Thus when analyzed the net benefit (as defined in the human benefit model) the result was a bit negative for the navigator during the first week or so. Fortunately, the negative benefit was more than compensated by the ability to carry out navigation actions smoothly later.

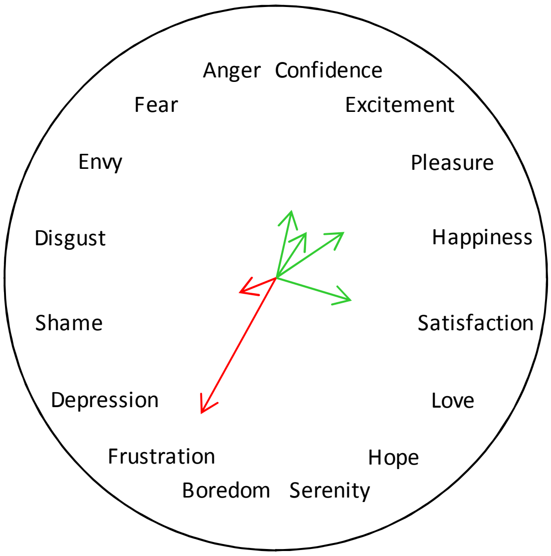

Figure U.3: Emotions created during the first experiences with a phone navigator.

For me a phone camera is a useful instrument when something unique happens unexpectedly. Thus, in my mind, the most important criterion for a phone camera is a fast and easy usage (rather than the number of megapixels). In that respect, the new camera exceeded my expectations. After taking about ten photos, I tested how fast I could take a photo when the phone was not in hand but was immediately available. The result was 15 seconds. It seems possible to automatize the usage of the camera in a way that a photo could be taken in less than 10 seconds.

Is it valuable to decrease the delay from 20 seconds to 10 seconds? This is an example in which time (or delay) has a special meaning. If something extraordinary happens, the difference between 10 seconds and 20 seconds might be essentially larger than the ordinary value of time for a person. The total waste of 10 seconds of time means about 0.10 Euros if we just consider the value of time. On the contrary, you may consider the value of missed photo being comparable to several Euros, or in very special situations much higher. An example of special situation is a view I saw some years ago: a snake swallowing a frog so that only the rear legs of the frog were visible. I missed the opportunity to take a curious photo because I did not have my phone available (I could have used the photo to illustrate the nature of ecosystems in this book). Because extraordinary situations are rare by definition, the expected benefit of a camera being available all the time is at most moderate.

During the first day with the new phone, I spent maybe four extra hours to get familiar with new features. Was that initial period beneficial enough to make the experience positive? In order to assess that experience separately from the mostly positive expectations of future benefits, I may consider a situation in which I only had this initial period of 24 hours without the possibility to use the phone (or any similar phone) later. Was the experience positive enough to bring about a positive net benefit?

Even more specifically, I could compare two options for spending those four hours: First, playing with the new phone, without the possibility to use it later. Secondly, writing something else about user experience, because that was what I planned to do.

Surely, the experience when playing with the phone exceeded the average experience during a typical writing period. However, an integral benefit of writing is to actually have some prepared text at the end of the process. If that completion benefit is added to the writing, then the two options appear equal to me. Then as to the later usage, a better display and higher bit rate significantly affected my web usage. As a rough estimate based on the first days of using the phone I spent three times more on surfing than earlier. After a couple of weeks, the usage decreased to some extent but still remained clearly above the level with my previous phone.

From where have I taken the extra time spent with the phone? Now I may spend somewhat less time with my home PC, reading newspapers, and watching TV. An appropriate mobile phone seems to offer an additional advantage when used for occasional news reading and watching. I may watch 1-minute TV news and check the weather forecast, and then put the phone away. It may take in total 3 minutes. If I had to do the same with an ordinary PC or TV, it typically takes much longer because it is so easy to keep on looking up other web-sites or other programs, etc. These kinds of behavioral changes are difficult to assess in advance, and in particular, they cannot be observed in a laboratory environment.

What is the reason for this difference in behavior between a phone and other media? First, although the quality of the display is high, the display is still small. Secondly, a high access rate consumes battery. Thirdly, my mobile service includes a monthly data limit. If the amount of data I have transmitted to my phone during a month exceeds the limit, the given bit rate drops. These are sacrifices I have to pay compared to an artificially perfect situation. Can restrictions and sacrifices then be beneficial? I would say, yes, sometimes they are. As an example, text messages cost something for the sender and the number of characters is limited. Because of these sacrifices the usage habits of text messages is essentially different from those of emails. The cost issues force the sender of text messages to consider, at least briefly, what to write and whether it is reasonable to send the message. The result seems to improve the quality of the interaction.

Measuring service quality

User experience is something we experience when we interact with a good or a service. Quality of user experience is a concept that is used to describe the essential part of the experience. However,

> **quality:** the inherent nature of an entity perceived by a human mind

is almost impossible to assess or measure reliably. It is even doubtful whether quality is one-dimensional concept. Although we often speak about higher and lower quality, are we ever able to describe the inherent nature of something within one dimension? From a modeling viewpoint, the answer is that several qualitative aspects shall be first combined (possibly with complicated functions and several weights) in a way that the result can be described on an ordinal, but still qualitative scale. As to the intuitive processes taking place in the mind, this combination process is complicated, hard to model and sensitive to external conditions.

Still, we are usually able to assess an experience on a scale consisting of qualitative attributes. Then a key objective of modeling user experience is to convert something measured on the quality scale to the linear benefit scale described in Chapter H.

It is also possible to apply a numerical scale, for instance, from 0 to 10. However, the end-points of numerical scale have to be defined by qualitative terms because the meanings of the end-points are not obvious (remember also the discussion about the eudemony scale in Chapter H). Moreover, there is no guarantee that different people define the middle of the scale in the same way.

Terms like excellent, good, and bad can be used in various situations because they do not refer to any particular aspect. Whatever terms are used they have to form an ordinal series in a way that they can be situated on a cardinal scale. For instance, we could assess the experience of a phone call in many dimensions: the service may create a soft or hard, a natural or artificial, and a normal or peculiar feeling. The opinions about the preferences may vary between participants and may depend on the properties of the situation.

ITU-T (International Telecommunication Union – Telecommunications Standardization Sector) has used Mean Opinion Score (MOS) scale to test the perceived quality of audio and video services (see Recommendation P.800, ITU-T, 1996). The scale is mainly used to assess the performance of different coding methods for voice and video applications. The scale is extended here because of the need to use as general models as possible. Most importantly, the extended scale includes score 0 which means a service that is totally useless. That level naturally corresponds to the zero-benefit level in the benefit model (see Figure H.8). It is also possible to consider situations in which the quality of a service is not just flawless but in which exceptionally high quality creates additional benefit per se. The extended scale is defined in Table U.2.

Table U.2: Extended MOS scale, scores 1-5 form the original MOS scale.

Score	*Quality*	*Impairment*	*Benefit*
6	Surprising	No impairment, quality is clearly better than what could be expected	Quality is able to create additional benefit
5	Excellent	Imperceptible	Expected benefit
4	Good	Perceptible but not annoying	
3	Fair	Slightly annoying	
2	Poor	Annoying	
1	Bad	Very annoying	Clearly reduced benefit
0	Useless	Total	No benefit at all

Now the modeling challenge is to convert MOS scores to a linear benefit scale. The approach adopted here is to make a questionnaire study in which participants have to make a choice between two options. In the first option, the quality of voice remains fixed during a 5-minute phone call. In the other option, the quality is changed during the call in a way that the user experiences two levels of quality. When assessing their preferences the participants use only the information shown in Table U.2 and their own intuition. Thus the results of the experiment primarily reflects the opinions about the terms *surprising*, *excellent*, *good*, *fair*, *poor*, *bad*, and *useless*. However, because the table also indicates numbers from 0 to 6, it can be assumed that they consider *fair* to be somehow in the middle between *poor* and *good*. In addition, the impairment descriptions can make the quality terms more concrete influencing also the assessment process.

The extensions to the original MOS scale, *surprising* and *useless*, are more problematic. First, it is difficult to find any term that describes quality that is clearly better than *excellent*. Thus, we need additional explanations. The basic idea here is to define the 6th level as a quality that is able to create additional benefit, for instance, in the form of pleasure: "I could never believe that the quality of a voice call could be as high as it is now" –type of experience. However, the trouble in this type of situation is that exceptional quality may actually disturb the fulfillment of the original objective of the voice call. Instead, the additional benefit of exceptional quality shall provide additional benefit even when the users are not paying any attention to the quality itself. It is certainly possible to increase the quality of typical voice calls in a way that makes it possible to perceive small but important nuances of voice.

Table U.3 shows the format of questions that can be used to assess the relative benefits created on different quality levels. The numbers of answers shown in the table are based on surveys made at Aalto University in 2010 - 2012.

Table U.3: Examples of questions to assess the benefit level on the extended MOS scale. Situation: your voice call lasts 5 minutes independent of perceived quality even when the quality is so low that the connection is totally useless. Question: which one of the two alternatives would you prefer?

Quality A	*prefer A*	*prefer B*	*Quality B*	
Excellent 5 min	30	42	*Surprising 4 min*	*+ Good 1 min*
Good 5 min	45	27	*Surprising 3 min*	*+ Fair 2 min*
Good 5 min	82	10	*Excellent 4 min*	*+ Poor 1 min*
Fair 5 min	71	21	*Excellent 4 min*	*+ Bad 1 min*
Fair 5 min	80	12	*Good 3 min*	*+ Poor 2 min*
Poor 5 min	17	55	*Excellent 4 min*	*+ Useless 1 min*
Poor 5 min	23	69	*Fair 4 min*	*+ Bad 1 min*
Bad 5 min	38	34	*Fair 3 min*	*+ Useless 2 min*

How can we then use the results to define the relationship between MOS-scale and a linear benefit scale? The major assumptions to perform the analysis are the following:

- Each MOS level corresponds to a fixed benefit level.

- The total benefit of the call is the integral of the momentary benefits.
- The participants of the survey try to maximize the total benefit of the call.

Based on these assumptions it is possible to define a simple function that is able to predict the answers to the questions in consistent way. Figure U.4 shows the result of evaluation based on 3089 individual comparisons.

The main expected observation is that the conversion between MOS-scale and linear benefit scale is non-linear: particularly the difference between poor and fair is much more significant than the difference between good and excellent. At least in case of voice service, something that is *good* is good enough and any improvement above good is marginal.

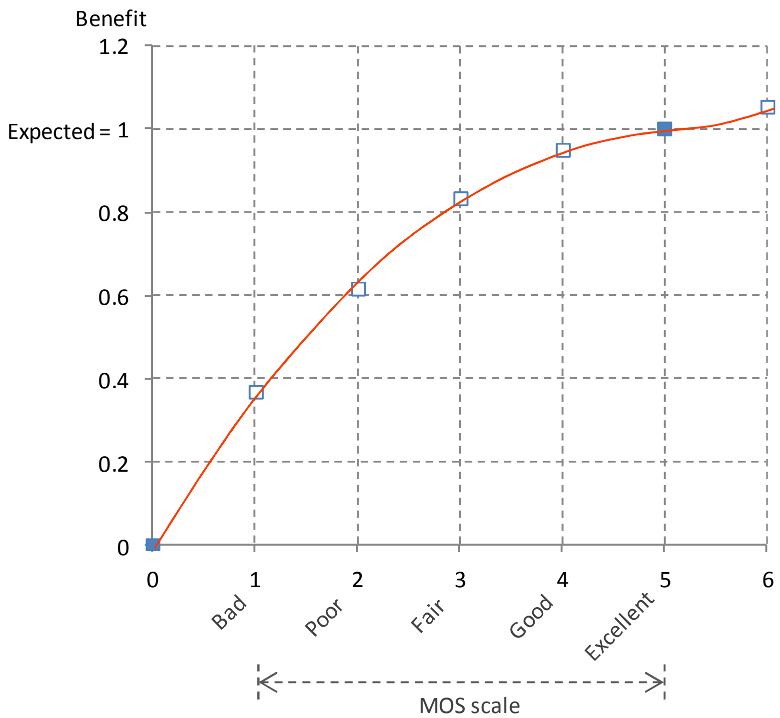

Figure U.4: Relationship between the extended MOS scale and linear benefit scale. Fixed points: excellent = 1 and useless = 0.

Another observations, although not as strong as the non-linear conversion, is that the benefit for poor quality seems to somewhat below what a smooth curve would predict. This deviation might be explained by the obviously small difference between attributes *poor* and *bad*. Because the comparison was made by using primarily the terms good, fair, etc., instead of the numeric scale or quality impairments, poor and bad might be difficult to distinguish from each other. This issue was further studied with 34 students that located 18 quality terms on a linear scale.

Although the mother language of most of the students was not English, the answers were highly correlated: the correlation between the average result and individual answers was typically between 0.95 and 0.99. The average position of the terms is presented in Figure U.5.

Indeed, there was no significant difference between poor and bad. Thus a more proper scale than the conventional MOS-scale would be:

5 = Excellent (or perfect)
4 = Good
3 = Ordinary
2 = Poor
1 = Awful (or useless)

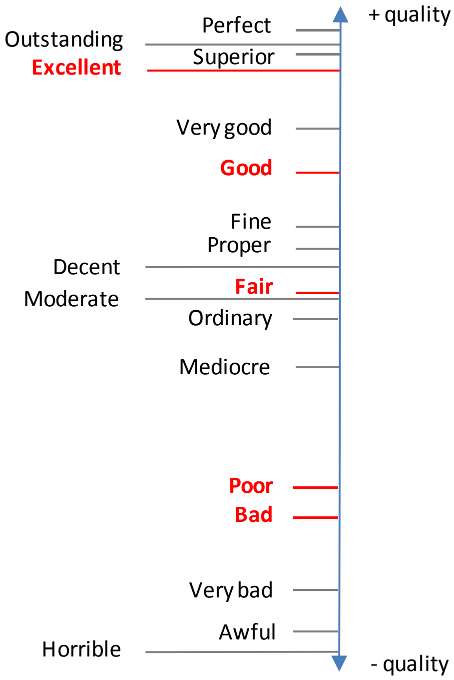

Figure U.5: Quality terms on a mentally linear scale based on the assessment of 34 students.

This is a generic problem of all survey studies. For instance, if a

```
Likert scale: a type of survey question where respondents are asked
to rate the level at which they agree or disagree with a given
statement
```

is used to assess the opinions about any matter, it is not at all clear that the offered choices (typically 5, 6, or 7) form a linear scale in the sense discussed above (see, e.g., Jamieson 2004).

Note also that a similar study made in a different context may result in an essentially different form than presented Figure U.4. For instance, if restaurants were assessed on a similar scale (perfect, good, ordinary, poor, and horrible) the difference between the value of good and perfect could be larger than in case of voice calls. The difference between these two cases is that in restaurants the quality of the cuisine is an integral part of the experience whereas with voice calls quality is almost irrelevant if a certain quality limit is exceeded (that is, good).

Modeling the benefit of data services

We could discuss incessantly about the experiences with new devices and applications and about the feelings they produce. Another matter is to assess the benefit of an application with a numerical model. This section offers an example how the general insight can be converted to a model. Let us consider a situation in which a user spends his time and mental effort to browse web pages using a mobile device. The experience depends on many issues, particularly on the content of the web pages, the speed of connection to the Internet, and the quality of the display.

Both the browsing application and the user are adaptive, that is, they react to the conditions of the network. The most essential task of the modeling effort is to take those adaptive properties properly into consideration. To simplify the analysis we assume that a session consists of two kinds of periods: downloading a web page and consuming the content of the downloaded page as illustrated in Figure U.6. In practice, the downloading of information and the usage of information may overlap. However, in this simplified model we ignore this phenomenon.

The downloading time depends on the speed of the connection (or access rate). The simplest assumption would be that the average size of a requested web page is independent of the access rate available during the session. An alternative assumption could be that the average downloading time for a requested web page remains constant. In the long term, that might be a reasonable assumption, because whenever the average access rate is increased, the average size of the content (measured in bytes) tends to be increased proportionally. For instance, when the access rate of customers is increased, content providers may add video clips that can consume any available excess capacity. In contrast, if we model the behavior of a user within a shorter time scale (e.g., one month), part of the increased access rate will be used to shorten the downloading time while part of the increased access rate will be used for downloading additional data.

Consequently, we may assume that the average amount of data downloaded per useful period depends on the access bit rate (R) as follows:

$$D_p(R) = D_p^* \left(\frac{R}{R^*}\right)^{\beta}$$

where R^* is the reference (or expected) access rate and D_p^* is the average amount of data downloaded for a period when the realized access rate is the same as the reference access rate. Parameter β describes the level of adjustment:

- If β = 0, then the amount of data remains constant, that is, users do not change their downloading behavior when the access rate is increased, but just exploit the decreasing waiting time.
- If β = 1, then the length of downloading time remains constant, that is, users exploit fully any increase of access rate by downloading more data.

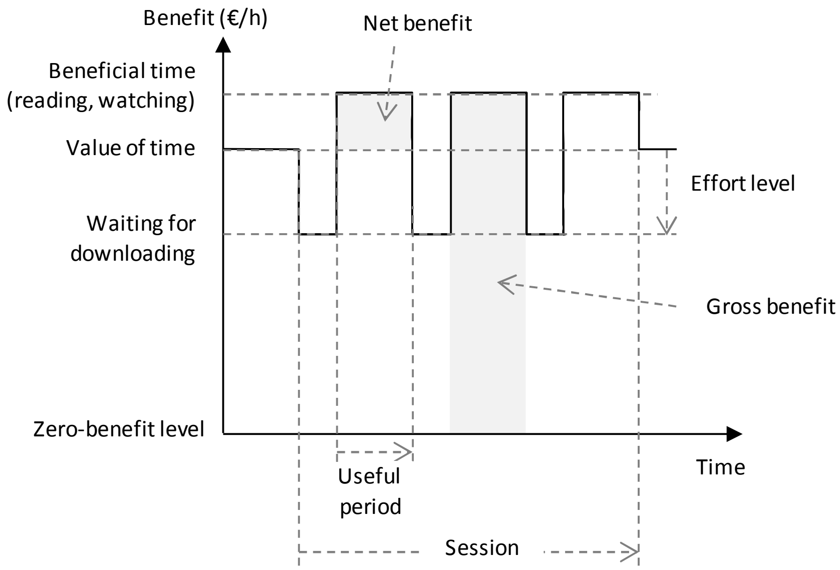

Figure U.6: A simplified model for a web browsing session.

Furthermore, there is another aspect that we need to consider: if a user is willing to download a larger amount of data when the access rate is increased, it means that he should get additional benefit from the increased amount of data. A possible choice to model this relationship is to use a logarithmic function:

$$B_p(D_p) = B_p^*\left(1 + \alpha\, ln\frac{D_p}{D_p^*}\right)$$

where D_p^* is the reference amount of data and B_p^* is the total gross benefit generated by the downloaded data.

In addition we need to make assumptions about the length of useful period (t_p) and the length of session (t_s). To avoid any excessive complexity we may assume that both of these lengths are independent of the amount of downloaded data. This assumption is somewhat questionable, because better quality due the higher amount of data will likely affect those time scales as well. However, we still have one more aspect of the model to consider, namely, the amount of sessions during a month. Therefore, although the lengths of useful period and session are kept fixed, the access bit rate will affect the amount of usage through the number of sessions. Note also that because we assume that the length of sessions remains constant, the average *number* of periods during a session is not constant but depends on the access bit rate.

Let us consider a case in which a user has a web browsing session by making the following assumptions:

- the reference access rate, R^* = 1 Mbit/s,
- the average length of session, t_s = 11 minutes,
- the average length of a useful period, t_p = 1 minute,
- the reference amount of data for a period, D_p^* = 750 kilobytes,
- the average gross benefit of a useful period, B_p^* = 0.75 Euros,
- α = 0.25,
- β = 0.50,
- value of time = 38 Euros per hour, and
- the effort level during waiting time = 0.4.

Based on these assumptions the average gross benefit of a session is:

$$10 \cdot \left((1 - 0.4) \cdot 38 \cdot \frac{6}{3600} + 0.75 \right) = 7.88\ €$$

because the downloading of 750 kilobytes takes 6 seconds, which means that in 11 minutes there will be 10 periods of useful time. Because the gross benefit of the session is somewhat above the value of time for the same duration (6.97 Euros), we may assume that the user will usually be satisfied with the sessions.

The main purpose of the model is to make predictions about user behavior when the access rate is changed. Table U.4 presents some results. The last column in the table shows the benefit factor that simply is the relative gross benefit per session compared to the reference case (1 Mbit/s in this example). Benefit factor is the main parameter within the overall framework of Ecosystem model, because it is used directly in the human benefit model (see Chapter H). The other parameter that is used in the ecosystem model is the average bit rate: it is the most critical parameter to define how much network resources are needed to provide good enough service quality.

In practice it is difficult to determine the values of different parameters (D_p^*, α, β, and so on). We have to remember that these parameters are mainly for internal purposes, with the aim to make the structure of the model comprehensible. For instance, although gross benefit per period (D_p^*) has only a minor effect on the benefit factor, it is still a necessary part of the model.

In this example, the only parameter that is "intentionally" changed is the access rate. We have to be, however, aware of the fact that whenever something is changed in a complex system, the change will inevitably affect many other things. Moreover, the changes force some other agents to make new decisions. If the access rate of a customer is upgraded, the network operator may increase the network capacity in order to keep the network performance and service quality intact. In contrast, if the operator does not upgrade the infrastructure at all, that likely means lower quality for other customers. As a result, the operator may need to react to the complaints of other customers.

Table U.4: Session characteristics as a function of access rate.

Access bit rate (Mbit/s)	*Data per period (kbytes)*	*Waiting time per period*	*Average bit rate (kbit/s)*	*Periods per session*	*Gross benefit per period (€)*	*Gross benefit per session (€)*	*Benefit factor*
0.1	237	18.9	24	8.4	0.53	5.47	0.69
0.5	530	8.5	62	9.6	0.69	7.12	0.90
1	750	6.0	91	10.0	0.75	7.88	1
2	1061	4.2	132	10.3	0.81	8.65	1.10
10	2371	1.9	307	10.7	0.97	10.43	1.32

Furthermore, after the access rate of a majority of customers is increased, there will certainly be an effect on the content offered by content providers. That process may actually *decrease* the benefit obtained by those customers that still use lower access rates. The approach promoted generally in this book and particularly in this specific case is to make:

> an **analysis:** a separation of a conceptual or material whole into its constituent parts and the study of the parts and their interrelationships.

The starting point is still the whole, while the separated parts primarily serve the understanding of the whole system. Thus, it is important to keep the model for each constituent part as simple as feasible in order to make it applicable in the more general framework. This also means that we need to define what phenomena are included in each separate model. For instance, the web-browsing model described above aims to model the effect of access rate to the behavior of individual users in a one month or less time frame. This particular model does not include the reactions of network operator, service provider, or content providers. Those phenomena shall be handled by other parts of the ecosystem model.

Network effect

What is a network effect? In the context of products and users it can defined as follows:

> **network effect:** the effect that one user of a product has on the value of the product to other users.

This is the user perspective. The other useful perspective is that of service provider. Typically, the "network effect laws" are presented from the service provider viewpoint. One of the standpoints of this book is that the driving force for the ecosystem as a whole is human benefit, thus the network effect is presented primarily from a human viewpoint. The difference is illustrated in Figure U.7. It might seem appropriate to take the service provider viewpoint immediately, because the objective usually is to assess the business opportunity rather than the benefits of users. However, the omission of the user benefit (at least as an intermediate step) may easily lead to unrealistic models. Anyway, we are, as human beings and users of various products, able to assess the usefulness of a product from the viewpoint of individual user. Any model aimed for evaluating real situations must also be reasonable from an individual viewpoint.

The inclusion of user benefit makes it also easier to assess situations in which users form different segments with different benefits, like innovators, early adopters, majority, and laggards. The network effect in each segment can be assessed separately, because the amount of benefit obtained by an innovator may depend in very different way on the amount of other users than the benefit obtained by a laggard. It might be that for an innovator the network effect is negative (because an innovator wants to be ahead of the majority) whereas laggards only perceive any benefit at all if a majority of others are already using the product.

Various formulas have been presented to model the network effect including the following ones presented here from the perspective of one user assuming that there are in total K identical users of a service:

- Sarnoff's law: $b_s(K) = c_s$.
- Logarithmic law: $b_l(K) = c_l log(K)$.
- Metcalfe's law: $b_m(K) = c_m K$.
- Reed's law: $b_r(K) = c_r 2^K / K$.

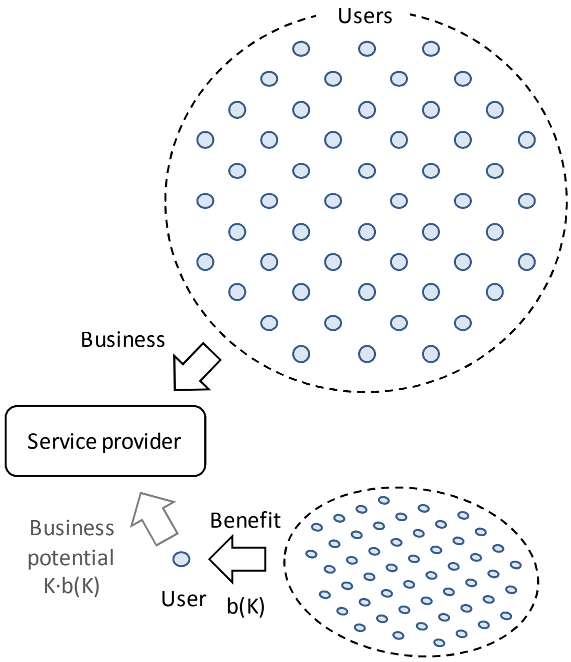

Figure U.7: Two viewpoints to network effect: business of service provider (above) and individual user (below).

The maximum business opportunity can then be simply assumed to be: $B_{max}(K) = Kb(K)$. Yet, it is not obvious that this simple relationship between user benefit and business opportunity is valid. However, if the stress is on the *opportunity* rather than on the business in a real situation, the assumption seems to be acceptable. Then as to the real business environment, many other factors affect the total business volume, particularly the competition between service providers and substitution effects between services. For instance, when the number of users increases, competition between providers may become tougher. Thus, the result might be that when the number of users increases the realizable business decreases as compared to the maximum business opportunity.

From user perspective, Sarnoff's law represents a situation in which there is no network effect at all. From service provider viewpoint, more customers apparently imply a larger business opportunity. As an example, if a user is reading news from web, the benefits do not directly depend on the number of other users. Nevertheless, the variety of news may still depend on the number of customers reading the news. If the business model is based on advertisements, the business volume depends linearly on the number of customers. As a result, a bigger business means more news and, hopefully, improved quality.

Only some internal features in the devices might generate benefits that are entirely independent of the number of other users of the same feature. For instance, the benefits of a camera in a mobile phone do not depend much on the number of cameras in other phones. The value of the camera in one phone may even be decreased when there are many other phone cameras available. Similarly, if you are the only person in your family with a navigator in your phone, the obtainable benefit might be considerable. Then if another member of the family also acquires a phone with navigator, the additional benefit is smaller.

Then if we consider a typical communications service, like phone calls or text messages, the amount of persons that you are able to call is of great importance. If we consider a community in which all people know each other, we may assume that the average benefit of adding one person to be fixed. As a result, we get Metcalfe's law. Note, however, that adding a new customer in one region of the country likely has a much smaller effect on the benefit for another person living in other part of the country. Thus, we need to be careful when using the seemingly valid formulation of Metcalf's law.

One popular concept some years ago was Reed's law (see http://en.wikipedia.org/wiki/Reeds_Law). Reed's law states that the value of a *K* member network is proportional to 2^K, because the total number of possible groups that can be constructed from *K* members is on the order of 2^K. The fundamental problem with Reed's law is the assumption that factor c_r describing the average value of a group remains constant even when *K* grows large. That is an unrealistic assumption. Let us assume that the total population were 1 000 000 persons and the total value of all possible groups for each person were 1000 Euros. Now if we remove 20 persons from the population, while 999 980 remains, the value of remaining groups would diminish down to 0.001 Euros per person. Although the example clearly reveals the problematic behavior of Reed's law as such, there still is certain wisdom in the Reed's basic idea. Still, a formal model has to be much more realistic.

The construction of the model described here consists of two parts: the effect of service penetration on the benefit of the service, and the effect of the size of the area that is covered by the service. Figure U.8 shows a case in which the service penetration is homogeneous around a person (in the figure, non-users are selected randomly).

From modeling viewpoint, the assumption of even likelihood of service usage is useful because it simplifies calculations. On the other hand, in reality there likely is a positive correlation between close people as regards the usage of a service. The circles around the person (marked by red dot) correspond roughly the preferred social group sizes (called also Dunbar numbers, see Zhou et al. 2005). It seems that humans are spontaneously organized in groups with certain sizes that obey a geometrical distribution.

The first level in the hierarchy consists of groups with 3 - 5 members. These groups can be called support cliques. On the second level, the size of the groups varies from 9 to 15 members. These groups are called sympathy groups. On the third level, the size of the group is 30 - 45 members. The largest group in which all members can personally know all other members is about 150 people. The hierarchy can be extended in a way that the next level consists of groups that are about 3.2 larger than the groups on the lower level.

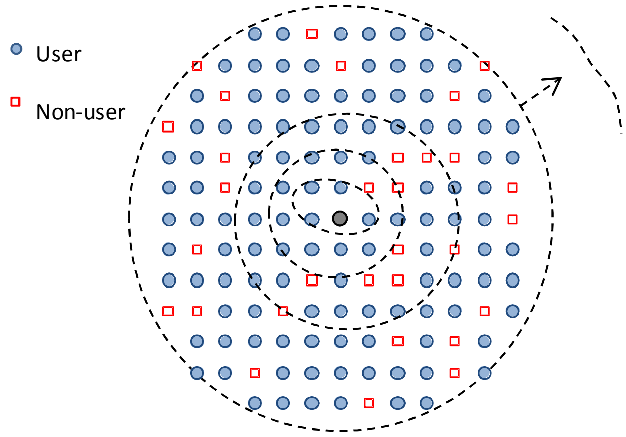

Figure U.8: Service penetration and circles of friends. Individual under study in the middle, circles represent roughly the sizes of social groups (Dunbar numbers).

If the size of group grows above 150 members, the group tends to split up into separate groups. Larger groups (e.g., with 500 members) can be kept together only through formal organization. It seems that we do not have, as human beings, sufficient capabilities to perceive and master the full complexity of groups above 150 members. Leaders with more than 150 subordinates most probably need to treat many people as a part of group rather than as individual people. Even if the leader may personally know more than 150 people, she cannot be aware of all significant relationships between her subordinates.

Consequently, it is unlikely that there are more than 150 people that are particularly important for a person. It might be nice to have several hundred people as a friend in a social media application. However, then the essential thing is likely the size of the number instead of the individuals that you truly want to communicate with.

Let us assume that the benefits of a networking service are grouped into three components:

1. Benefits that do not require any other persons, like remote access of information services.
2. Benefits that are obtained when connecting with another person, like telephone calls.
3. Benefits that are obtained when a group of persons uses the same service, like the usage of text message as an information channel for a team.

Let us first consider the benefit of a service for one individual person in a large population of M inhabitants. We may assume that there is a limit for the value of each service component for the person as follows:

b_1 = the total benefit of everything not depending on the service penetration,

b_2 = the total benefit of connecting two persons, and
b_3 = the total benefit of making connections among groups of persons.

The unit of each component could be Euros or dollars per month. If the penetration of the networking service is p = K/M, the average value of the whole service for an individual is

$$b(p) = b_1 + b_2 p + b_3 f(p)$$

where $f(0) = 0$, $f(1) = 1$ and $f'(p) \geq 0$. What could be a reasonable form for this part of the benefit function? The approach here is to first consider the total benefit of all groups of size *i*, and denote it by $B_g(i)$. The total value of all groups for a person is

$$\sum_{i=3}^{\infty} B_g(i) = b_3$$

where $B_g(i)$ denotes the total value of all groups of size *i* for a person.

By making appropriate independency assumptions, we can calculate the probability that all other members of a group of size *i* are using the service as follows:

$$q(i, p) = p^{i-1}.$$

Note that because the person to be considered belongs to the active persons by definition, the question is whether all other persons in a group are also using the service.

Here we also assume that the service is useful only if all members of the group are using the service. For instance, if a sport team wants to arrange their daily information transfer by text messages, even one missing person from the distribution list significantly decreases the usefulness of the service.

Moreover, to make the formula practical we need to assess the total benefit provided by different sizes of group. Although we might, in principle, assess separately the importance of all possible groups, in practice we need a simpler and more practical approach. Here we assume that the total value of groups of size *i* obeys the following geometrical distribution:

$$B_g(i) = c\,\alpha^{i-3} \;\; for\, i = 3, 4, 5 \ldots, \; otherwise\, B_g(i) = 0.$$

This assumption leads to the following result (see Kilkki and Kalervo 2004, for more detailed discussion):

$$B(p, r) = \frac{p^2}{r - 2 - (r - 3)p}$$

where *r* is the average size of groups weighted by the benefit each group produces:

$$r = \frac{\sum_{i=3}^{K} i\, B_g(i)}{\sum_{i=3}^{K} B_g(i)} \approx \frac{3 - 2\alpha}{1 - \alpha}.$$

Thus from the viewpoint of an individual user, the benefit of the service as a function of service penetration is:

$$b_g(p) = b_1 + b_2 p + b_3 \frac{p^2}{r-2-(r-3)p}\,. \qquad (U.1)$$

The business opportunity of a networking service can be obtained simply by multiplying $b_g(p)$ by the total number of persons using the service ($K = p \cdot M$). How does this model relate to the sizes of social groups? If we assume that the smallest groups (with 3 - 5 members) are more important than the larger groups, the result is that parameter *r* in KK-law is approximately the same as the size of groups on the second level of hierarchy, that is, r varies from 9 to 15.

This concluded the first part of the analysis of the network effect. The other part of the analysis concerns the size of the region covered by the service. Note particularly that if two service providers cover the same region a crucial issue is whether any seamless interoperation occurs between the providers. With perfect interoperability, the service providers divide the business opportunity proportionally by the share of customers, because the total service penetration is the sum of service penetrations of the two service providers. In contrast, without appropriate interoperability the penetration (*p*) in Formula U.1 is the service penetration of each service provider. As a result, without cooperation the third part in Formula U.1 will practically disappear. Consider, for instance, a situation in which you could send text messages only to other customers of the same service provider.

We can safely assume that a great majority of benefits provided by any communications services is obtained within the group of people one knows personally (including maybe 150 persons). Actually, a majority of social interaction, such as voice calls and text messages, occurs with a few people. Some of those people may live in distant locations, which makes it somewhat difficult to assess the effect of geographical coverage on the benefits of the service. Odlyzko and Tilly (2006) have argued that logarithmic function is reasonable way to model the effect of the total size of a customer base.

If we consider a very wide area with a lot of variations in service penetration, it is difficult to assess what should be used as (effective) service penetration. However, if we simply make the assumption of constant penetration we may apply the following formula:

$$b_g(p,K) = c\,log(K)\,\left(b_1 + b_2 p + b_3 \frac{p^2}{r-2-(r-3)p}\right). \qquad (U.2)$$

Figure U.9 illustrates the behavior of Formula U.2 in two cases with different weights for b_1, b_2, and b_3.

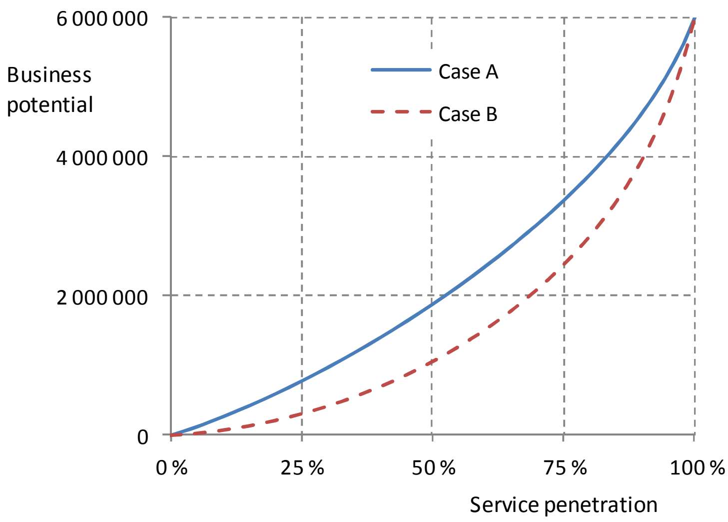

Figure U.9: Business potential as function of service penetration.
Case A: M = 1 000 000, $c = 1$, $b_1 = 0.5$, $b_2 = 0.3$, $b_3 = 0.2$, and $r = 10$.
Case B: M = 1 000 000, $c = 1$, $b_1 = 0.1$, $b_2 = 0.5$, $b_3 = 0.4$, and $r = 10$.

What is the relevance of this analysis for a CEE? It will be difficult to estimate the values of parameters, especially before the service is widely used. For instance, hardly anyone was able to predict the variety and amount of text message usage before the service was available. In retrospect, it is easier to describe what has happened. Case B in Figure U.9 gives a basis for a judgment. Some types of usage do not need any other users, such as receiving configuration data to the phone ($b_1 = 0.1$). A majority of text messages are delivered between two persons, often as a series of messages ($b_2 = 0.5$). What makes the text message service really valuable is that it can be used to serve large groups of people with one channel ($b_3 = 0.4$). If the service penetration falls below 95 percent, much of the gains of general availability are lost. Under these assumptions, the business potential with 100 percent penetration is almost 6 times larger than the business potential with 50 percent service penetration.

Internet connectivity is another example of a service with considerable network effect. Case A in Figure U.9 illustrates the situation. A substantial part of web-usage is independent of the number of other users ($b_1 = 0.5$). Yet, email and other personal communication tools are important for most of us ($b_2 = 0.3$). Finally, many web applications are much more useful when most of your friends are using them ($b_3 = 0.2$). The network effect is significant but still smaller than in the case of text messages.

Can we infer something from this formula, or is it useful just for illustration purposes? An interesting phenomenon is what happens when the customer bases of two service providers are combined or two networks are interconnected. Let us consider an example in which there are two service providers serving different regions, for instance, adjacent countries. The results of a brief analysis are presented in Table U.5.

Table U.5: Interconnection of text message services provided by two service providers. M(Small) = 1 000 000, M(Big) = 10 000 000, $c = 1$, $b_1 = 0.1$, $b_2 = 0.5$, $b_3 = 0.4$, and $r = 10$.

	Business potential (× 1000)					
Service penetration	*Separate service*		*Interconnected service*		*Interconnection gain*	
Small - Big	*Provider Small*	*Provider Big*	*Provider Small*	*Provider Big*	*Provider Small*	*Provider Big*
50% - 50%	1 061	12 467	1 254	12 545	194	77
75% - 75%	2 453	28 711	2 888	28 884	435	173
100% - 50%	6 000	12 467	6 290	12 863	290	396
100% - 75%	6 000	28 711	6 518	29 119	518	408
100% - 100%	6 000	70 000	7 041	70 413	1 041	414

What are the main lessons we can learn? Typically, the smaller provider seems to gain more than the bigger provider does, because each customer of the smaller provider gains significantly more than what a customer of the bigger provider gains. In contrast, if the service penetration of the larger service provider is significantly lower than the service penetration of the smaller provider, the balance is changed: the larger provider may even benefit more than the smaller one. This is a significant issue when the network effect is strong (that is, when both b_3 and r are large). Furthermore, it seems reasonable for a bigger provider to pay extra when they want to cooperate with a smaller provider if the smaller one has significantly higher service penetration in their own region (row 100% - 50% in Table U.5). In contrast, if the providers have approximately the same service penetration, the bigger provider has much stronger bargaining power. However, the ratio is not the same as the ratio in size (10:1) but approximately 2.5:1 under the assumptions made in this analysis.

In all cases, both providers have a strong incentive to increase the service penetration in their own region. For instance, if we take as a starting point the case "75% - 75%" in Table U.5, the bigger provider gains as much when it is able to increase its own service penetration from 75 percent to 75.2 percent as it gains from the interoperability with the smaller provider. However, that case only considers the immediate effects; cooperation may still provide significant positive effects in the long term for the bigger provider.

Book recommendations

M. Hassenzahl, 2010, *Experience Design, Technology for All the Right Reasons,* Morgan & Claypool.

In this concise book, Marc Hassenzahl clarifies the meaning of experience in the design of new devices and applications. Engineers are inclined to stress the practical usefulness of new products and features at the cost of emotional benefits and sacrifices. In order to avoid this kind of bias, the designers shall systematically consider the psychological needs of users. Hassenzahl's informative book can be greatly helpful in that task.

V. Kaptelinin and B. A. Nardi, 2006, *Acting with Technology*, *Activity Theory and Interaction Design*, Cambridge, MA: The MIT Press.

Activity theory provides a systematic way to consider the nature of human activities. In particular, it is important for the designers to take into account the wider context of all interactions and avoid concentrating in individual operations. The book provides a theoretical perspective that helps the designers of new products to fulfill the needs of customers.

D. A. Norman, 2011, *Living with Complexity*, Cambridge, MA: The MIT Press.

Donald A. Norman is the leading author in the area of usability and product design. This book provides invaluable insight for anyone who wants to understand the basic dilemma of current communications products: how enormous complexity can be made understandable for the users of the products.

References

Jamieson, S., 2004, Likert scales: how to (ab)use them, Medical Education, 38: 1217–1218.

Kilkki, K., M. Kalervo, 2004, KK-law for group forming services, Proceedings of XVth International Symposium on Services and Local Access, Edinburgh, UK, available at http://kilkki.net/3.

Odlyzko, A., B. Tilly, 2006, A refutation of Metcalfe's Law and a better estimate for the value of networks and network interconnections, available at http://www.dtc.umn.edu/~odlyzko/.

Rust, R. T., D.V. Thompson, R.W. Hamilton, 2006, Defeating Feature Fatigue, Harvard Business Review 84(2): 98–107.

ITU-T, 1996, Methods for subjective determination of transmission quality, Recommendation P.800, available at http://www.itu.int/itu-t/recommendations/index.aspx?ser=P

Zhou, W.-X., D. Sornette, R. A. Hill, R. I. M. Dunbar, 2005, Discrete hierarchical organization of social group sizes, Proceedings of the Royal Society B, 272: 439–444.