Stakeholder Logistics of an Interactive System

Johan F. Hoorn
Vrije Universiteit
De Boelelaan 1081a
1081 HV Amsterdam, NL
jfhoorn@few.vu.nl

ABSTRACT

In dealing with several key concepts in HCI, task analysis, and requirements engineering, I hypothesize that the Usability of interactive systems depends on their Effectiveness, Efficiency, and Effort. In its turn, Usability is one of the predictors of Satisfaction. Effectiveness relates to the degree of goal achievement with the system. Efficiency is based on the speed and accuracy of process execution. Effort has to do with the amount of energy invested in handling or understanding the system. Stakeholders supposedly require the maximum degree of goal achievement, as fast as possible and at the smallest possible error rate, against an optimal level of effort. Features of the machine and of the human side of an interactive system are assessed for their Relevance and Valence towards these requirements. Complex evaluations can occur if the judgments about the machine do not run parallel to those about the human operator. Several claims are corroborated by preliminary evidence from survey data sampled from 1943 employees of a multinational bank in the Netherlands, who interacted with 25 different banking systems.

Categories & Subject Descriptors

H.1.2 [Models and Principles]: *User/Machine Systems–Human information processing*; K.6.3 [Management of Computing and Information Systems]: *Software Management–Software development*.

General Terms

Requirements engineering, Human Factors, Theory.

Keywords

Stakeholder goals, effectiveness, efficiency, effort, performance, usability, satisfaction.

1. INTRODUCTION

Task analysis aims at identifying the work processes and procedures involved in operating interactive systems. Investigating stakeholder goals is crucial to understand the different stages of task execution and for designing appropriate computer aids. Task analysis can help to determine the time duration and possible error sources of performing a task. Moreover, it is useful for signaling the not-so-easy steps in the process of task execution. Task analysis is not necessarily restricted to the workfloor but can also include the decomposition of business processes or the effects of organizational culture on work [Clegg1993] [Anderson1994]. In doing task analysis, system design can rely on a cognitive model of stakeholder (i.e. user) performance, which may improve the usability of and satisfaction with an interactive system.

Although at face value, information and task analysts will admit that general requirements on an interactive system concern, among others, the system's effectiveness, efficiency, and demand on effort, these terms are not unequivocally used, often blend together during analysis, while their relationship with usability and satisfaction is under debate.

In this paper, I attempt to make an analysis of several key concepts in HCI, task analysis, and requirements engineering, integrate them into a unifying model, and make this model open to empirical scrutiny. In spite of their rather different demands [Lamsweerde2004], my claim is that different stakeholder groups share at least three goals, which pertain to effectiveness, efficiency, and effort, the intricacy of which should be the core concern of any task analysis or user-centered system design.

2. STAKEHOLDER LOGISTICS

Whether they are system administrators, maintenance personnel, business managers, or end-users, these and many other stakeholder groups deal with the supply, distribution, and exchange of information in an interactive system. These 'stakeholder logistics' feed the implicit and explicit judgments of how effective, efficient, difficult, or easy the system is or ought to be.

Barnard [1938] states that effectiveness indicates in how far the actual outputs of a system correspond to the desired outputs. That is, in how far a stakeholder achieves a desired goal with the system [ISO 9241-11, 1998]. Efficiency, for quite a few authors, refers to time aspects of process execution or task completion (e.g., [Frøkjær et al. 2000], [ISO 9241-11, 1998]). However, others claim that speed per se is meaningless unless it is combined with levels of accuracy [Käki 2004], [UlrichHebert 1982]. Effort relates to the amount of labor stakeholders put into handling a system [Cooper1968], [Silverstein et al. 1998] but can also refer to cognitive load [Oviatt et al. 2004].

Love [1991] states that effectiveness and efficiency are two crucial aspects of performance (also [Nielsen, 1994], Jordan [1998]). It seems likely that effort is the third because in particular efficiency is negatively affected by an increase in task-difficulty [Oviatt et al. 2004]. Yet, although it is widely acknowledged that performance as reflected by effectiveness and efficiency contributes considerably to usability (e.g., ISO 9241-11 [1998], [Nielsen 1994]), effort is mentioned only indirectly (cf. the 3-click rule) as a predictor of usability [Cooper1968].

The status of satisfaction is somewhat awkward in this respect. Authors such as [Brooke et al. 1990], [ISO 9241-11, 1998], and [Frøkjær et al. 2000] conjecture that satisfaction explains usability. Simply put, these authors claim that what makes you happy is more usable. In my view, this inference suffers from a fallacy in logical intentions. All things that make a person happy are not *necessarily* useful things but all useful things *possibly* make people happy. Satisfaction is also used as an established means of assessing the effectiveness of information systems [Pather et al. 2003]. Contrariwise, [Scott 1995] reviews literature claiming the opposite that effectiveness of an information system is responsible for user satisfaction. In all, I would argue that satisfaction is elicited *after* goal achievement, whereas usability and use relate to task execution while achieving such goals. Satisfaction, then, is the end-product of interacting with a usable system and not a stage in between (Figure 1).

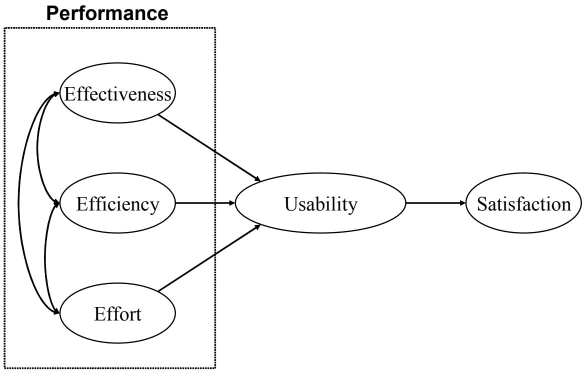

Figure 1. Model of Stakeholder Logistics. Usability of an interactive system is a mediator between Satisfaction on the one hand and Performance (Effectiveness, Efficiency, and Effort) on the other.

In Figure 1, Usability is modeled as a threefold multiple regression function of Effectiveness, Efficiency, and Effort. The three Performance factors are modeled as three independent variables. [Frøkjær et al. 2000] found in a correlational study that effectiveness, efficiency, usability, and satisfaction do not significantly correlate. In following ISO 9241-11 [1998], these authors posit that effectiveness and efficiency are directly responsible for the level of usability of a system. Satisfaction is the third but, as said, I believe satisfaction is a more general judgment than usability. Love [1991] also states that effectiveness and efficiency are independent variables (Section 4).

Other authors, however, argue that (a subset of) the three Performance factors are dependent. [Oviatt et al. 2004], for example, manipulated the levels of task difficulty and found that the number of task-critical errors and response latencies increased significantly when the task became more difficult. Thus, it seems that efficiency as a function of speed and accuracy is related to the degree of cognitive processing load or effort. In Figure 1, then, the three Performance factors are mutually correlated (double-headed arrows). This indicates, for example, that an increase in Effort can be negatively related to Efficiency or that an increase in Efficiency is positively related to high Effectiveness.

The assumptions illustrated by Figure 1 are corroborated by survey data gathered from 1943 employees of a multinational bank in the Netherlands (Appendix 1). Different from the findings by [Frøkjær et al. 2000], in that banking company, Satisfaction with the interactive systems depended for 16% on Usability, whereas Usability depended on Satisfaction for no more than 4%. Moreover, as predicted, Efficiency and Effectiveness explained 34% of the variance of Usability (regrettably, there were no questions on Effort in this survey). In support of the model in Figure 1, Efficiency and Effectiveness were positively correlated ($r = .53$) but could still independently explain part of the variance of Usability (Appendix 1).

Nonetheless, the 16% contribution to Satisfaction shows that Usability may not be the only satisfier in a system. For instance, [Shackel 1991, p. 25] defines usability not only in terms of effectiveness but also as learnability, flexibility, and attitude. [Jordan 1998] defines usability not only as efficiency but also as guessability, learnability, and re-usability. Stakeholder Logistics (Figure 1) merely states that the contribution of Usability to Satisfaction is pertinent and significant. All other factors that explain Satisfaction fall outside the scope of this model.

In the next sections, I will elaborate on the three Performance factors and discuss the concepts of Effectiveness, Efficiency, and Effort for both their machine and their human side. After all, both humans and machines are making up the ‘interactive system’ and both parties can contribute to making errors, being quick, and achieving goals. Both sides anchor in perceptions of the physical and directly countable variables of a system, such as number of target hits or clock time. This means that if a system is considered ‘fast,’ this judgment is based on the perception of the physical clock time to execute a task to which the stakeholder, moreover, attributes a subjective weight. The hypothesis I will try to maintain is that (H1) stakeholders require the largest possible degree of achieving a personal or business goal with an interactive system (effectiveness). They want this as fast as possible against the smallest possible number of errors (efficiency) and at an optimal (which need not be minimal) level of effort.

3. EFFECTIVENESS (OUTPUT)

In following Drucker [1954][1974], Love [1991] asserts that effectiveness is related to goal end states (“doing the right things”), whereas efficiency is process oriented (“doing things right”). [Frøkjær et al. 2000] envision effectiveness as “the outcome of the user’s interaction with the system.” Interpreted within Fitts’ [1954] paradigm, effectiveness would be a reflection of the number of times the user hits a target in contrast to the number of misses. However, judgments on effectiveness are not bound to low-level interaction issues but can also concern achieving business goals and other higher objectives with a system (e.g., [HamiltonCervany1981]). In that case, effectiveness can also be a function of hitting and missing a number of business targets.

Effectiveness of an interactive system, then, is related to the degree that a stakeholder achieves a goal with the system. This goal may be related to the computer task at hand (e.g., to find a word in a document) or something outside that particular computer task (e.g., being entertained by the animated agent of a help function). Stakeholders assess effectiveness in terms of result [Love 1991] and success [Seddon et al. 1998]. Seddon et al. [1998] and Novick [1997] make an important distinction between the effectiveness of humans and the effectiveness of machines in goal accomplishment. Effectiveness can be measured by the degree that a stakeholder perceives or experiences that a goal is achieved (whether by the machine or by themselves). Envisioning the stakeholder logistics as an assembly line, effectiveness is the result of successfully putting together the end-product (cf. Barnard [1938]).

and that of the machine, a positive correlation is hypothesized, pointing out that an increase in effectiveness of the machine probably co-occurs with higher effectiveness of the human and v.v.

Finally, the judgment of human effectiveness and the judgment of machine effectiveness are combined into one overall judgment of the Effectiveness (effective vs. ineffective) of the interactive system. In its turn, interactive Effectiveness explains part of the variance of Usability, which explains part of the variance of Satisfaction (Figure 1).

Taking a closer look at Figure 2 reveals that between the physically countable variables and experience, the effects of two moderators occur: Relevance and Valence [HoornVeer 2003a, 2003b] [Hoorn et al. 2005] after [Frijda 1986]. That is, stakeholders estimate the relative importance (Relevance) of the physical variable to their personal goals and concerns and estimate in how far this damages or supports those goals and concerns (Valence). Only then can experience take place and do emotions occur [Frijda 1986]. A more detailed account of the Relevance-Valence Moderator Box is provided in Section 6. In the next section, I will argue that Effectiveness is relatively dependent on but not similar to Efficiency.

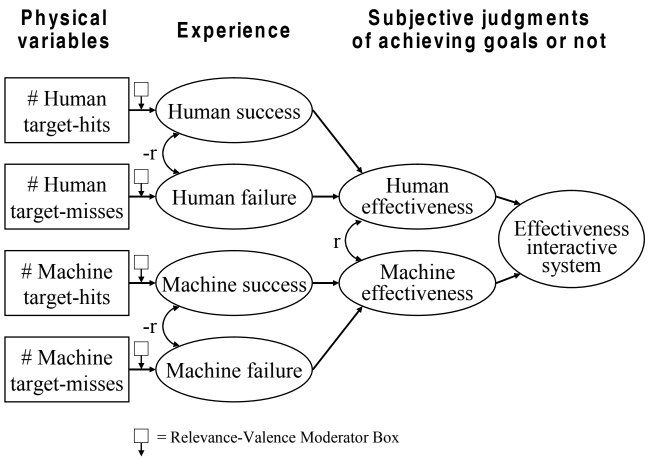

Figure 2. Explanation of judgments on effectiveness of an interactive system.

In Figure 2, Effectiveness of an interactive system is modeled not as an aspect inherent in a system but as a judgment (whether by experts or not) about that system. These judgments are fed by physically countable variables, such as the number of target hits and misses by both human and machine. Stakeholders experience the relative number of hits and misses in terms of success and failure, respectively. Common sense would have it that the relation between the numbers of hits versus misses is expressed by a ratio or percent. However, from a stakeholder's point of view, this is not necessarily the case. It is not hard to imagine that certain stakeholders are positively biased to their own success and underestimate their own degree of failure. In addition, power users may attribute machine success to their own doings and their own human failure to the machine. Novices or low-achievers may do the opposite, thinking they are too clumsy to handle a complex machine. In other words, stakeholders put different weights or relevance to the number of hits and misses so that the relation between hits and misses can differ from conventional ratio. Therefore, the experience of success and failure are modeled in Figure 2 as two relatively dependent variables; their interrelation being expressed by a negative correlation (-r) (Figure 2, double-headed arrows).

Figure 2 further indicates that the experience of success and failure precedes the more formal and reflective judgments of how effective (effective vs. ineffective) the human aspect of interaction was in hindsight. The same occurs for the machine aspect (effective vs. ineffective). Between human effectiveness

4. EFFICIENCY (THROUGHPUT)

Although in the literature, the definition of effectiveness is somewhat confused, compared to efficiency the discussion seems to be transparent. In quite a few cases, aspects of effectiveness are attributed to efficiency and the other way round or efficiency is treated as a mere aspect of time.

To start with the ISO 9241-11 [1998] norm, effectiveness is defined as the accuracy and completeness of goal achievement. ISO regards error rates as one of the indicators of effectiveness (also [Nielsen et al. 1994]) and I suppose as related to the accuracy aspect. Efficiency, so the ISO norm runs, pertains to the *relation* between accuracy and completeness of goal achievement. Moreover, time to execute a task is regarded a stable indicator of efficiency (also [Nielsen et al. 1994]). [Frøkjær et al. 2000] follow the ISO definition of efficiency in focusing on the time aspect alone.

In view of Section 3, there will be not much discussion that completeness of goal achievement is the core of effectiveness. However, accuracy of goal achievement seems to be mistaken. A goal can be completely achieved without being achieved very accurately (just being lucky). A goal can be completely missed although the accuracy of the shot was perfect (bad luck). In other words, accuracy seems more like an aspect of the process towards goal achievement than an aspect of goal achievement in itself. Put differently, "it is quite possible for a manager to work efficiently and still remain ineffective" [Wambugu1982]. In support of that, Love [1991] states that "even the most efficient organization cannot survive if it is efficient at doing the wrong things. Likewise, the organization with the greatest effectiveness can disintegrate from poor efficiency."

Thus, if accuracy does not belong to goal achievement but to process execution then it is not an aspect of effectiveness but rather of efficiency and so are the related error rates. What remains then from the ISO definition of efficiency is the time aspect of process execution. Yet, as ISO 9241-11 [1998] states, there can be a *relation* between accuracy (of the process) and goal achievement and I will return to this matter later.

From the previous paragraphs it can be learned that efficiency relates to task or process execution and that it has an accuracy and a time aspect. Efficiency is subordinate to effectiveness because it is just a means to achieve a goal. The experiential interpretation of the time to execute a process or to complete a task is speed ("My system is fast, your system is slow"), which is relative to previous experiences. The Usability Glossary [2005] states:

When asked to perform a task as well as possible, people will apply various strategies that may optimize speed, optimize accuracy, or combine the two. For this reason, comparing the performance of 2 users cannot be done on the basis of speed or accuracy alone, but both values need to be known [Usability Glossary 2005].

To be meaningful, hence, speed should be taken together with accuracy (i.e. error rates) for humans [Käki 2004] as well as for machines [UlrichHebert 1982]. In other words, process execution is governed by the well-known phenomenon that increases in speed are traded for decreases in accuracy and v.v [Usability Glossary 2005] [Oviatt et al. 2004]. Efficiency then is one of the states of a speed-accuracy trade-off that a process can be in. This can apply to perceptual-motor and other human tasks as well as to running a process on a machine.

Thus, efficiency of an interactive system has little to do with the goal in itself as it is related to the means or processes to achieve those goals. The relation between accuracy and goal achievement that ISO 9241-11 [1998] hinges on should be seen as follows. Achieving a desired goal or not may depend on the accuracy of the associated process. In a way, accuracy has to do with achieving certain sub goals that lead to achieving the main goal. However, not every sub goal has to be reached in order to get to the main goal. In other words, there may be a relation between accuracy and the completeness of goal achievement but the first is not necessary to achieve the other (e.g., [Wambugu1982]).

Efficiency should be decomposed into a time (i.e. speed) aspect and an accuracy aspect. The real-time a process takes to achieving a goal with the system (e.g., connecting to the Internet) can be experienced as fast or slow. Time-experience, then, is one of two components of efficiency. Accuracy, whether the real number of errors humans and machines make is experienced as error prone or not, is the other. If connecting to the Internet is fast but at the cost of making many errors, the process is deemed – at least partially – inefficient. If the user accurately follows a wizard but at the cost of being extremely slow, the wizard also seems (partially) inefficient.

My position is that the degree of estimated efficiency is a function of the experience of two speed-accuracy trade-offs that each can be in one of four states (Figure 3). Process execution by humans [Käki2004] or machines [UlrichHebert1982] can be experienced as slow and inaccurate, slow but accurate, fast but inaccurate, or fast and accurate. The latter combination is usually considered the optimal state of a system's processes but does not need to be in all cases. For example, in their meta-analysis Nielsen et al. [1994] found that the highest speeds in process execution did not necessarily lead to the highest levels of satisfaction. Using the analogy of the assembly line, efficiency is a function of the speed of the assembly belt and the precision with which, for instance, all the different parts of a computer chip are mounted on a surface.

Figure 3 displays the constellation of speed-accuracy trade-offs in process execution of both humans and machines as based on clock time and precision. Time is a continuous variable and therefore one would expect that the experience of process time is either "fast" or "slow." Figure 3, however, shows that the same time epoch of a (stage in the) process can be experienced as fast as well as slow. This has to do with the interference of Relevance and Valence (Section 6). Suppose connecting to the Internet from a remote location is urgent (Relevance is high) to check an important message in the e-mailbox. Suppose this message tells the reader whether s/he is accepted for a tenure track position or not. Then the outcome Valence of the message can be positive (acceptance) or negative (rejection). If the applicant anticipates acceptance, connecting to the Internet will be appraised as slow rather than fast ("I can't wait"). Conversely, if the applicant fears to be rejected, the same time to establish the connection is experienced as (too) fast. If the applicant is in doubt about the outcome of the message, parallel experiences of fast as well as slow will occur. This evaluative inconsistency becomes even clearer when the time to reboot after a failure (machine experienced as slow) equals the time a virus needs to erase a disk (too soon).

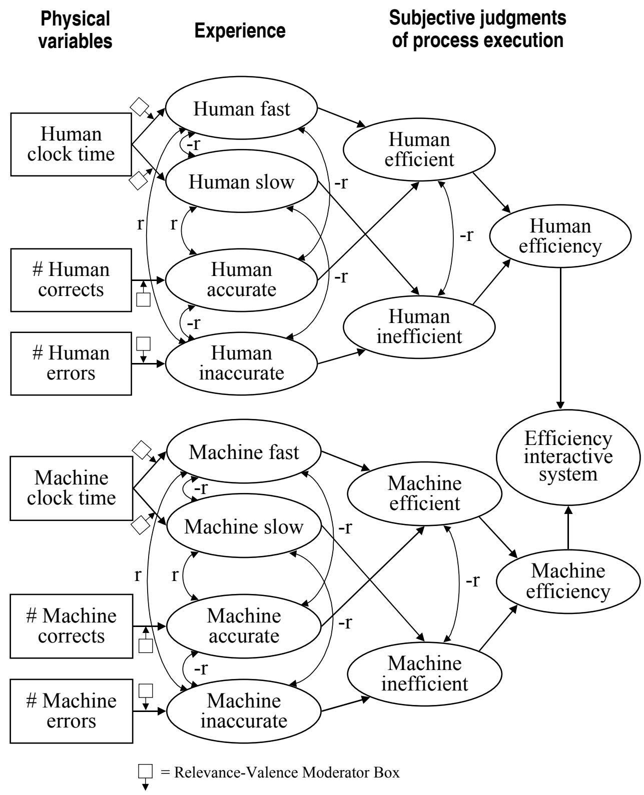

Figure 3. Explanation of judgments on efficiency of an interactive system.

The other two physical variables that foster efficiency judgments are the number of correctly executed stages and procedures in a process in some sort of negative relation (-r) to the number of errors. This situation resembles the one described for the numbers of target hits and misses (Section 3). The relation between corrects and errors logically would be a percentage but psychologically, estimated accuracy may not be fully related to estimated inaccuracy.

The speed-accuracy trade-offs that occur at the level of physical variables are reflected in the stakeholder's experience as a set of four correlations for humans and four for machines (Figure 3). That a process is appraised as fast is positively related to the experience of inaccuracy (faster and with more imprecision) and negatively to accuracy (faster and with less precision). That a process is experienced as slow is positively related to accuracy (slower and with more precision) and negatively to inaccuracy (slower and with less imprecision). Experiences of "fast" in combination with "accurate" lead to the judgment of efficiency. Experiences of "slow" in combination with "inaccurate" lead to the judgment of inefficiency. Mixed combinations are also possible. Processes that are regarded fast as well as inaccurate (e.g., quick and dirty methods) will be partially considered efficient as well as partially inefficient. Processes that are regarded slow as well as accurate (e.g., check-and-double-check security policies) will show a similar ambivalence in the efficiency judgments. Subsequently, the combined judgments on human and machine efficiency and inefficiency predicate the level of Efficiency (efficient vs. inefficient) that is attributed to the interactive system as a whole. This judgment explains the degree of Usability, which is responsible for a significant part of the variance in Satisfaction (Figure 1).

Again, I turned to the survey data sampled from 1943 employees of a multinational bank in the Netherlands (Appendix 1). Definitely, the assumption was corroborated that Efficiency judgments were based on experiences of the Speed and Accuracy of an interactive system. Multiple regression analysis showed that Speed and Accuracy significantly explained 21% of Efficiency. In addition, the proposed decomposition of Speed and Accuracy into Fast vs. Slow and Accurate vs. Inaccurate indeed was fruitful: Slow and Inaccurate significantly predicted 11% of the Inefficient score. However, Fast and Accurate could not significantly explain the variability of the Efficiency score, an inconsistency that at this point cannot readily be explained.

5. EFFORT (INPUT)

In discussions on effectiveness, efficiency, and usability of interactive systems, effort seems to be out of the picture – although not completely.

In Barnard's [1938] vision on efficiency, certain aspects of effort are mixed in with aspects of effectiveness. Barnard characterizes efficiency as the ratio of completed outputs to actual inputs, such as money, skills, or workforce. ISO 9241-11 [1998] follows the same line of thought in that efficiency is not only the "accuracy and completeness with which users achieve certain goals" but also the "resources expended in achieving them."

In my view, outputs are typical for goal achievement (effectiveness) whereas money, skills, workforce, and other resources are more related to the amount of effort (in a broad sense) [Eason 1988] that is invested into the process. In a computer task, such resources could be cognitive capabilities invested in the speed-accuracy trade-off during task execution [Oviatt et al. 2004] [Käki 2004]. Comparing inputs (e.g., effort) to outputs (i.e. the number of completed products) is more of a measure of cost-effectiveness or a return-upon-investment estimation than a description of efficient throughput of materials or information *during process execution*.

Bevan et al. [1991] state that usability is largely dependent on the ease of use of a product both ergonomically and as mental effort. They rely on the ISO standard for software qualities [ISO 1991], which says that usability is "a set of attributes of software which bear on the effort needed for use and on the individual assessment of such use ..." They cite [Eason 1988] in that ease-of-use is "the degree to which users are able to use the system with the skills, knowledge, stereotypes and experience they can bring to bear."

Effort, then, is not an aspect of effectiveness and efficiency because in itself, effort is goal and process independent. Although effort can be related to, for instance, efficiency [Oviatt et al. 2004], a goal can be achieved without any effort (being lucky) or not achieved despite the effort (bad luck). A process can be efficient although it takes a lot of energy to execute it (e.g., air transport) or a process can be inefficient although the work it takes is not hard (e.g., doing many small work-arounds). Effort is the individual experience of the energy it takes to execute a process so to achieve a certain goal. This effort could be real work power or the symbolic representation of such power, such as money. In many cases, stakeholders want this effort to be as little as possible (easy to use, understandable, comfortable, minor workload, etc.). Particularly in game applications, however, the effort it takes to master the skills is highly appreciated. In other words, the effort is not only negative and something to be avoided but is considered optimal [Yerkes-Dodson1908] according to a subjective criterion value of the stakeholder. The default level of acceptable effort can be determined by the genre of the application (e.g., game vs. tutorial). In terms of the assembly line, effort is the amount of energy the engine takes to put and keep the assembly belt into gear.

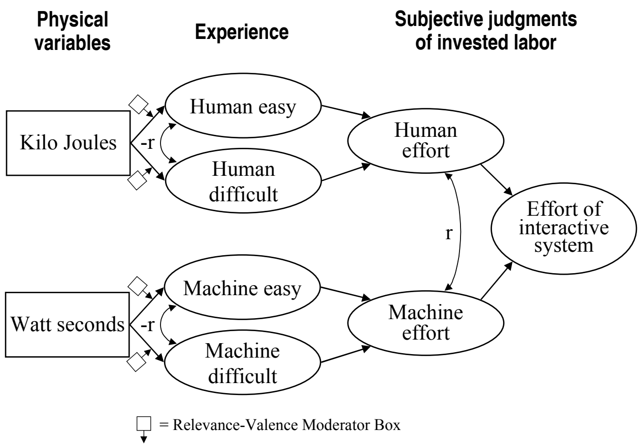

Figure 4. Explanation of judgments on effort expended in working with an interactive system.

In Figure 4, two indicators of human and machine energy consumption are depicted, KJ and Wsec, respectively. Yet, these may not be the only physical variables that could be filled in. Money, number of employees, barrels of oil, or other indicators of resource depletion may do just as well. Here also, although the physical countable variables lie on a continuum, the experience of such variables is discontinuous – that is opposite evaluations can occur in parallel. For instance, if an important (Relevance) and profitable (positive Valence) e-business transaction takes a lot of effort to complete, the effort is still deemed less compared to the same amount of effort invested in an irrelevant business transaction with negative outcome expectancies ("it doesn't matter to us and they won't bite anyway"). In other words, if humans or machines have a hard time in executing a process or achieving a goal, the job may be evaluated as difficult, mixes of difficult and easy, or perhaps even easy, dependent on the levels of Relevance and Valence (Section 6). Difficulty and ease stand in a negative relation (-r) to one another, which does not mean that an increase in difficulty is related to the same amount of decrease in ease and v.v.

Human effort and machine effort stand in a positive relation to one another. If the machine has an easy job in task execution, the human will feel that his/her workload is less than if the machine has a tough job. Together, human effort and machine effort determine the overall judgment of Effort invested in an interactive system.

As mentioned, effort has an optimum (inverted-U curve, cf. [Yerkes-Dodson1908]), which need not be minimal. In sum, then, Usability of an interactive system is a curvilinear function of Effort, together with a linear function of Effectiveness, and a linear function of Efficiency (Figure 1).

6. RELEVANCE AND VALENCE

In Hoorn et al. [2005], empirical evidence is reported that Valence has a moderating effect on agreement to requirements on an interactive system under development. The effects of Relevance were not investigated in that study but for simplicity, I assume that Relevance also acts as a moderator. That is, after the physically countable variables have been encoded by the stakeholder, these are compared with the three main requirements that people have on an interactive system: A sufficient number of target hits (effectiveness), the right combination of speed and accuracy during process execution (efficiency), and an optimal level of effort. With regard to efficiency, Figure 5 has filled in the speed-accuracy combination with minimal time against minimal error because this is the state that is preferred most of the time.

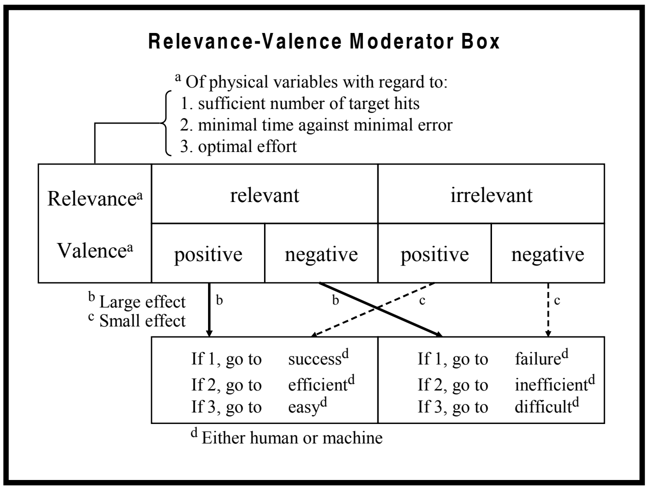

Figure 5. Physical variables are evaluated on Relevance of and Valence towards the three main requirements people have on an interactive system.

Features of the human-machine system that are considered relevant to (one of) the three main requirements evoke more intense reactions (Figure 5, drawn arrows – large effects) than features that are judged irrelevant (Figure 5, dashed arrows – small effects). For example, if connecting to the Internet through broadband triples processing speed, broadband is a relevant feature. An aesthetic GUI – although pleasurable in itself – is an irrelevant feature with respect to the requirement of efficiency. Reactions to establishing or interrupting the connection through broadband, then, are more intense than a change in the appearance of the dialog boxes during process execution.

Features are also evaluated for their potential to facilitate or inhibit (Valence) the three main requirements. If an increase in Internet processing speed is most wanted, the outcome expectancy with regard to broadband will be positive and with regard to an analog telephone line, it will be negative. If the attitude towards a feature is positive, that feature contributes (Figure 5) to the level of success (Figure 2), efficiency (Figure 3), or ease (Figure 4) of the system. If the attitude is negative, the feature contributes (Figure 5) to the level of failure (Figure 2), inefficiency (Figure 3), or difficulty (Figure 4).

7. CONCLUSIONS/DISCUSSION

In this paper, an attempt has been made to improve the conceptual precision of several important notions in HCI, task analysis, and requirements engineering. The model of Stakeholder Logistics was introduced to better assess the performance of an interactive system in terms of a combination of effectiveness, efficiency, and effort of both humans and machines. It was argued that three main requirements on an interactive system pertain to maximum goal achievement (effectiveness), the highest possible speed against the highest possible accuracy during task or process execution (efficiency), at the cost of an optimal level of effort.

As an important result, models have been developed in which concepts such as performance, usability, and satisfaction have found their place. The triplet effectiveness, efficiency, and effort have been made more precise – from physical foundation to psychological experience – and their mutual relationships have been pointed out.

Moreover, the distinction between the human side of the system and the machine aspects have made it possible to understand sometimes-contradictory judgments about the usability of an interactive system. It may well be that a machine is inefficient in executing a task but due to the clever work-arounds of the user, the overall performance of the system may seem to be alright.

The addition of the ‘emotion’ variables relevance and valence has made it possible to explain the different experiences of the same physical variables. Dependent on the level of importance and the positive or negative outcome expectancy, the time to achieve a desired goal, for instance, is experienced as longer than the same amount of time that is needed to achieve an undesired goal. Thus, improving a system on the physical level (e.g., faster processing time, more precise algorithms) may not automatically lead to higher levels of experienced efficiency or more satisfaction on the side of the stakeholder.

The models proposed in this paper have been made open to empirical verification. This can be done, for example, by means of structured questionnaires or reaction time recordings (i.e. regarding the speed-accuracy trade-offs during interaction). Several claims were empirically verified using data from four years of IT satisfaction research at a multinational banking company. The hypotheses were corroborated that usability explained satisfaction more than the other way round. Moreover, usability depended on at least effectiveness and efficiency, the latter two being partially correlated. Further, efficiency of the bank’s IT depended on measures of speed and accuracy, as predicted. Stakeholders that deemed the IT inefficient thought so because the processes were regarded as slow and inaccurate. However, no significant evidence was found that IT that was considered efficient depended on experiences of processes being fast and accurate.

That the latter relation remained absent may be a problem of the data set. The data were part of the organization’s internal evaluation of the IT and IT use. They were not gathered for theoretical purposes. Moreover, the data were sampled with single items, which make it impossible to do reliability analysis of the measurements. In other words, the (absence of) empirical evidence presented in this paper should be considered as provisional. Yet, these data provide firmer ground for theoretical considerations than intuition alone.

To gather more reliable data, I am currently investigating the business processes in the operation room (OR) of an academic medical center to develop an electronic efficiency monitor. Based on this electronic score board, the throughput of patients and OR allocation time is supposed to be optimized. In parallel, I will work with the national training center for air traffic control to predict the performance (effectiveness, efficiency, and effort) of pupils while interacting with their monitoring equipment during special maneuvers such as ‘approach’ and ‘airplane landing.’

8. ACKNOWLEDGMENTS

This paper is part of the Innovation Oriented research Program (IOP) for Human-Machine Interaction entitled *Integrating Design of Business Processes and Task Analysis,* granted to the author by the SenterNovem Agency of the Dutch Ministry of Economic Affairs in The Hague, grant Mmi99009. Johan F. Hoorn works in the Faculty of Sciences, Dept. of Computer Science, Section Information Management & Software Engineering. Brenda Neuteboom of the ICT IDM Management of the bank is kindly thanked for making the data available.

9. REFERENCES

[1] Anderson, R. 1994. Representations and requirements: The value of ethnography in system design. Human-Computer Interaction 9: 151--182.

[2] Barnard, C. I. 1938. The Functions of the Executive. Harvard University Press, Cambridge, MA.

[3] Bevan, N., Kirakowski, J., Maissel, J. 1991. What is usability? In Human Aspects in Computing: Design and Use of Interactive Systems and Work with Terminals, Elsevier Science.

[4] Brooke, J., Bevan, N., Brigham, F., Harker, S., and Youmans, D. 1990. Usability statements and standardisation - work in progress in ISO. In Proceedings of the IFIP TC13 Third International Conference on Human-Computer Interaction, 357--361

[5] Clegg, C. 1993. Social systems that marginalize the psychological and organizational aspects of information technology. Behaviour and Information Technology 12: 261--166.

[6] Cooper, W. S. 1968. Expected search length: A single measure of retrieval effectiveness based on the weak

ordering action of retrieval systems. Journal of American Society of Information Science, 19 (1): 30-41.

[7] Drucker, P. F. 1954. The Practice of Management. Harper and Row, New York, NY.

[8] Drucker, P. F. 1974. Management, Tasks, Responsibilities, Practices. Harper and Row, New York, NY.

[9] Eason, K. 1988. Information technology and organisational change. Taylor and Francis, Bristol, PA.

[10] Fitts, P. 1954. The information capacity of the human motor system in controlling the amplitude of movement. Journal of Experimental Psychological*,* 47: 381--391.

[11] Frøkjær, E., Hertzum, M., Hornbæk. K. 2000. Measuring usability: Are effectiveness, efficiency, and satisfaction really correlated? In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, The Hague, NL, 345--352.

[12] Hamilton, S. and Chervany, N. 1981. Evaluating information systems effectiveness – part 1: Comparing evaluation approaches. MIS Quarterly 5 (3): 55--69.

[13] Hoorn, J. F. 2005. ITO 01+02+03+04, Technical Report. Amsterdam, NL, Computer Science, IMSE.

[14] Hoorn, J. F. and Van der Veer, G. C. 2003a. Requirements analysis and task design in a dynamic environment. In Human-centred Computing: Cognitive, Social, and Ergonomic Aspects. Volume 3 of the Proceedings of HCI International 2003, Hersonissos, Greece, 472--476.

[15] Hoorn, J. F. and Van der Veer, G. C. 2003b. Requirements engineering for groupware to manage business process dynamics and stakeholders' mindset. In Proceedings of the 9th European Conference on Cognitive Science Approaches to Process Control: Cognition and Collaboration - Distributed Cognition in Complex Processes, Amsterdam, NL, 153--160.

[16] ISO 1991. Software Product Evaluation - Quality Characteristics and Guidelines for their Use. ISO DIS 9126.

[17] ISO 1998. Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs) - Part 11: Guidance on Usability. ISO 9241-11.

[18] Jordan, P. W. 1998. An Introduction to Usability. Taylor and Francis, London, UK.

[19] Käki, M. 2004. Proportional search interface usability measures. In Proceedings of the third Nordic conference on Human-computer interaction, Tampere, Finland, 365--372.

[20] Love, A. J. 1991. Internal Evaluation: Building Organizations From Within. Sage, Newbury Park, CA.

[21] MacCallum, R. C., Zhang, S., Preacher, K. J., and Rucker, D. D. 2002. On the practice of dichotomization of quantitative variables. Psychological Methods 7 (1): 19--40.

[22] Nielsen, J. and Levy, J. 1994. Measuring usability: Preference vs. performance. Communications of the ACM 37 (4): 66--75.

[23] Novick, D. G. 1997. What is Effectiveness? Position paper for the CHI '97 Workshop on HCI Research and Practice Agenda Based on Human Needs and Social Responsibility, Atlanta, GA, 33--36.

[24] Oviatt, S., Coulston, R., and Lunsford R. 2004. When do we interact multimodally?: Cognitive load and multimodal communication patterns. In Proceedings of the 6th International Conference on Multimodal Interfaces, State College, PA, 129--136.

[25] Pather, S., Erwin, G., and Remenyi, D. 2003. Measuring e-commerce effectiveness: A conceptual model. In Proceedings of the 2003 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on Enablement through Technology, Pretoria, SA, 143--152.

[26] Scott, J. E. 1995. The measurement of information systems effectiveness: Evaluating a measuring instrument. ACM SIGMIS Database 26 (1): 43--61.

[27] Seddon, P. B., Staples, D. S., Patnayakuni, R., and Bowtell, M. J. 1998. The IS effectiveness matrix: The importance of stakeholder and system in measuring IS success. In Proceedings of the International Conference on Information Systems, Helsinki, Finland, 165--176.

[28] Shackel, B. 1991. Usability – context, framework, definition, design and evaluation. In Human Factors for Informatics Usability. Cambridge, UK, 21--37.

[29] Shackel, B. (1991). Usability-Context, framework, definition, design and evaluation. In: B. Shackel & S. J. Richardson (Eds.), Human factors for informatics usability, Cambridge: Cambridge University, 21-37.

[30] Silverstein, C., Henzinger, M., Marais, J., and Moricz, M. 1998. Analysis of a Very Large Alta Vista Query Log. Technical Report 1998-014. Palo Alto, CA, COMPAQ Systems Research Center.

[31] Usability Glossary 2005. Speed-accuracy Tradeoff. Retrieved March 09, 2005 from http://www.usabilityfirst.com/ glossary/term_475.txl

[32] Wambugu, C. K. 1982. The library manager. In Reader of the Unesco International Symposium on the Harmonization of Education and Training Programmes in Information Science, Karfa, Nairobi, chapter 10.

[33] Yerkes R. M. and Dodson, J. D. 1908. The relation of strength of stimulus to rapidity of habit-formation. Journal of Comparative Neurology and Psychology 18: 459--482.

10. APPENDIX 1

10.1 Usability-Satisfaction Study

I hypothesized (body text, Figure 1) that Satisfaction with an interactive system depends on its Usability, which in its turn depends on Effectiveness, Efficiency, and Effort. The latter three variables are assumed to be partially correlated. Other authors, however, state the opposite that Satisfaction would explain Usability.

10.2 Method

10.2.1 Participants, systems, and procedure

To verify these claims, I was allowed to use the data sampled from four subsequent years (2001-04) of nation-wide user satisfaction research by an ICT department of a multinational bank in the Netherlands [Hoorn2005]. In total, 1943 employees from different departments (e.g., Bank Shops, Advice Offices, Fulfillment, Sales Management, and Credit Support) participated in a yearly conducted electronic query (ITO 2001-04) about 25 interactive systems used in the organization (e.g., Signature Authorization, I-forms/HTML-forms, Card/PIN Activation, Business Transactions, etc.). This ITO survey consisted of 195 single items and open questions of which a subset of 75 items was useful for the present purposes. These 75 items pertained to Satisfaction, Usability, Effectiveness, and Efficiency. Unfortunately, no items on Effort were included in the list.

10.2.2 Measurements

Satisfaction was measured by 25 items on interactive banking systems, which were rated on a ten point scale (1= bad, 10= excellent). Not every bank employee used every system that is possibly available. A worker in the bank shop may operate the Mortgage Advice Program but not Peoplesoft Rollout Complaints. To yet obtain an overall measure of Satisfaction with the available interactive systems, I calculated for each employee an average Satisfaction with the system(s) s/he did use (grand mean average Satisfaction, M= 6.49, SD= 1.06).

Usability of the systems was measured by the item "The PC and systems that I work with are sufficiently user friendly to do my work properly," rated on a 6-point scale (1= completely disagree, 6= completely agree) (M= 3.74, SD= 1.20). Effectiveness was supposed to be indicated by the item "The PC and systems that I work with provide sufficient information to do my work properly,"[1] also rated on a 6-point scale (M= 4.38, SD= 1.04). Efficiency was supposed to be indicated by the item "The PC and systems that I work with allow me to quickly and adequately accommodate my client's questions and needs," rated for agreement on a 6-point scale (M= 3.67, SD= 1.34).

10.2.3 Regression analysis 1

To investigate whether Satisfaction with a set of 25 interactive systems depended on their Usability or the other way round, I conducted two multiple linear regression analyses (method Enter). Because not every bank employee used each system, excluding missing data case wise reduced the sample size to N= 928. The first regression analysis simulated the conception depicted in Figure 6. That is, Usability was supposed to explain Satisfaction, whereas Effectiveness and Efficiency were included as controls.

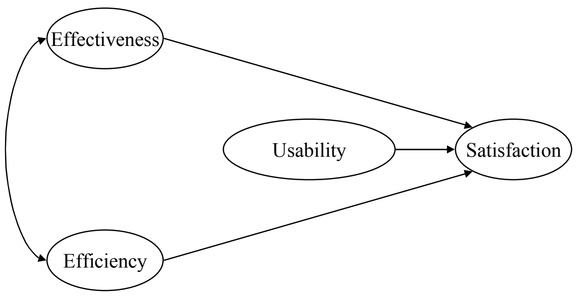

Figure 6. Usability explains Satisfaction, whereas Effectiveness and Efficiency serve as controls.

In regression 1 (method Enter), the dependent variable was Mean Satisfaction with Interactive Systems and the independent in step 1 was Usability of PC and Systems. The independents in step 2 were Information provided by PC and Systems, indicating Effectiveness, and Working quick and adequately with PC and Systems, indicating Efficiency.

10.2.4 Results regression analysis 1

Usability accounted for a significant 16% of the mean Satisfaction variability, R^2= .16, R^2_{adj}= .16, $F(1,926)$= 181.85, p= .000. Effectiveness and Efficiency incremented the explained variance of Satisfaction with only 4%, although this contribution was significant, R^2_{change}= .04, $F_{change}(2,924)$= 25.83, p= .000. On the basis of correlation-regression analyses, the relative importance of Usability, Effectiveness, and Efficiency in predicting Satisfaction was assessed. It seemed that Usability was most strongly related to mean Satisfaction, standardized β= .26, t= 7.19, p= .000. Supporting this conclusion is the height of the standardized Beta coefficient and the strength of the positive correlation between Usability and Satisfaction partialling out the effects of all other predictors ($r_{partial}$= .23, r_{part}= .21). Efficiency was the second best predictor (standardized β= .21, t= 5.76, p= .000, $r_{partial}$= .19, r_{part}= .17), whereas Effectiveness offered little or no additional predictive power ($p > .05$) beyond that contributed by Usability and Efficiency (this finding counters [Scott1995] but is nevertheless not a confirmation of [Pather et al. 2003]).

10.2.5 Regression analysis 2

The second regression analysis followed the set up exposed in Figure 7. Effectiveness and Efficiency should predict Usability, whereas Satisfaction served as a control.

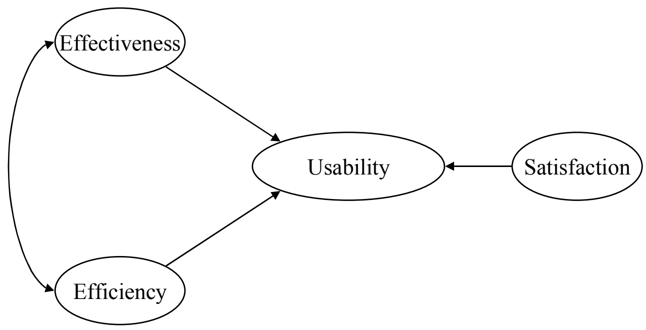

Figure 7. Effectiveness and Efficiency explain Usability, whereas Satisfaction serves as a control.

In regression 2 (method Enter), the dependent was Usability of PC and Systems. Independents in step 1 were Information provided by PC and Systems (i.e. Effectiveness) and Working quick and adequately with PC and Systems (i.e. Efficiency).

---

[1] After all, getting information is the main goal of working with an information system.

The independent variable in step 2 was Mean Satisfaction with Interactive Systems.

10.2.6 Results regression analysis 2

Efficiency and Effectiveness together accounted for a significant 34% of the Usability variability, R^2= .34, R^2_{adj}= .34, F(1,925)= 236.58, p= .000. Mean Satisfaction incremented the explained variance of Usability with 4%, R^2_{change}= .04, F_{change}(1,924)= 51.62, p= .000. The relative importance of Effectiveness, Efficiency, and Satisfaction in explaining Usability showed that Efficiency was most strongly related to Usability, standardized β= .35, t= 10.15, p= .000. Supporting this conclusion is the height of the standardized Beta coefficient and the strength of the positive correlation between Efficiency and Usability partialling out the effects of all other predictors ($r_{partial}$= .32, r_{part}= .26). Effectiveness was the second best predictor (standardized β= .25, t= 7.96, p= .000, $r_{partial}$= .25, r_{part}= .21). Although close to Effectiveness, mean Satisfaction was the last in line (standardized β= .21, t= 7.18, p= .000, $r_{partial}$= .23, r_{part}= .19).

10.2.7 Conclusions/Discussion

Satisfaction with an interactive system depends more on its Usability (16%) (R^2= .16, regression 1) than the other way round (4%) (R^2_{change}= .04, regression 2). Yet, the contribution of Satisfaction to Usability is significant, which may indicate that after people are satisfied with a system, their default level of estimated usability of that system is increased when they start using that system for the second or the third time. Put differently, when the process of using the system starts all over again.

In addition, Usability is significantly predicted by Efficiency and Effectiveness. Together, they explain 34% of the variance of Usability (R^2= .34), whereas Satisfaction only contributes 4% (see previous paragraph). Efficiency and Effectiveness showed significant correlations (r= .53) but they could yet predict part of the variance of Usability in a relatively independent way (β= .35, $r_{partial}$= .32, r_{part}= .26 for Efficiency and β= .25, $r_{partial}$= .25, r_{part}= .21 for Effectiveness).

In all, these results seem to support the assumptions depicted by Figure 1 (body text) that Usability is a mediator between Satisfaction on the one hand and Effectiveness and Efficiency on the other. The latter two are relatively independent, which is manifested by the highest significant correlations and yet independently maintaining explanatory power.

10.3 Efficiency Study

Next, I will explore some of the assumptions on Efficiency put forth in Section 4, (Figure 3, body text). There I assumed that Efficiency judgments are based on experiences of the Speed and Accuracy of process execution. In an interactive system, such judgements can apply to the human influence on the system and/or to the machine's influence.

10.4 Method

10.4.1 Measurements

To measure Efficiency, in Section 10.2.2 I used the item "The PC and systems that I work with allow me to quickly and adequately accommodate my client's questions and needs," rated for agreement on a 6-point scale (M= 3.67, SD= 1.34). To see whether this item could be explained by experiences of speed and accuracy, I selected two more items from the ITO 2001-04 survey list that explicitly mentioned time and error aspects encountered in the task environment. Speed was indicated by "How satisfied are you with the waiting time at the Helpdesk telephone line" (M= 4.42, SD= 1.26). Accuracy supposedly was keyed by "The PC and systems that I work with are so error free that I can do my job properly," (M= 2.86, SD= 1.26).

10.4.2 Regression analysis 3

In regression analysis 3 (method Enter), the dependent variable was the Efficiency item on working quick and adequately with PC and Systems. The independents were the Speed item about the waiting time at the Helpdesk telephone and the Accuracy item about PC and Systems being error free.

10.4.3 Results of regression analysis 3

The Speed and Accuracy item accounted for a significant 21% of the Efficiency variability, R^2= .21, R^2_{adj}= .21, F(1,758)= 101.12, p= .000. Accuracy was most strongly related to Efficiency, standardized β= .43, t= 13.22, p= .000. Supporting this conclusion is the height of the standardized Beta coefficient and the strength of the correlation between Accuracy and Efficiency partialling out the effects of all other predictors ($r_{partial}$= .43, r_{part}= .43). On top of that, Speed also contributed independently to Efficiency (standardized β= .10, t= 2.94, p= .003, $r_{partial}$= .11, r_{part}= .10). These results support the conclusion that Efficiency indeed is a function of Speed and relatively independent of that, a function of Accuracy.

10.4.4 Regression analyses 4 and 5

In Section 4, body text, I assumed that experiences of 'fast' and 'accurate' together explain judgments of interactive systems being 'efficient,' whereas 'slow' and 'inaccurate' would feed judgments of systems being 'inefficient.' To arrive at different groups of bank employees that would represent the judgments efficient vs. inefficient, fast vs. slow, and accurate vs. inaccurate, I performed a median split for the Efficiency item (Median= 4.00, Cum%= 68.3), Speed item (Median= 4.00, Cum%= 46.9), and Accuracy item (Median= 3.00, Cum%= 68.4).[2] The employees who judged that the Efficiency was high thus established the score for Efficient; those who thought Efficiency was low established the score for Inefficient. Likewise for the Speed item, which was divided into a score for Fast and a score for Slow as well as the Accuracy item, which was divided into a score for Accurate and a score for Inaccurate.

To investigate whether judgments of efficiency depended on experiences of 'fast' and 'accurate' and inefficiency depended on 'slow' and 'inaccurate,' I performed two stepwise multiple regressions (method Enter). In regression analysis 4, the Efficient score served as dependent, the Fast and Accurate scores being the predictors, and Slow and Inaccurate being the controls. In regression analysis 5, the Inefficient score served as dependent, the Slow and Inaccurate score being the predictors, and Fast and Accurate being the controls.

10.4.5 Results of regression analyses 4 and 5

The model assessed with regression 4 was insignificant (F < 1). However, the model assessed in regression 5 did hold. Slow and Inaccurate accounted for a significant 11% of the Inefficient score variability, R^2= .11, R^2_{adj}= .10, F(2,251)= 14.82, p= .000.

[2] Although certain authors [MacCallum et al. 2002] oppose to the median split procedure as a loss of information, in this case, there was not any other means to verify hypotheses of unipolarity.

The Inaccurate score was most strongly related to the Inefficient score, standardized β= .28, t= 4.61, p= .000, i.e. when partialling out the effects of all other predictors ($r_{partial}$= .28, r_{part}= .28). In addition, Slow also contributed independently to the Inefficient score (standardized β= .14, t= 2.28, p= .023, $r_{partial}$= .14, r_{part}= .14).

10.4.6 Conclusions/Discussion

The data from the selected ITO 2001-04 survey items indeed supported the assumption that Efficiency judgments are based on experiences of the Speed and Accuracy of process execution (Section 4, Figure 3, body text). Regression analysis 3 showed that the Speed item and Accuracy item significantly accounted for 21% of the variability in agreement to the Efficiency item. Moreover, it seems worthwhile to decompose Speed and Accuracy into experiences of Fast vs. Slow and Accurate vs. Inaccurate, respectively. As expected by the theory (Section 4, Figure 3, body text), the Slow score and the Inaccurate score significantly contributed (11%) to the Inefficient score. Alas, no evidence was obtained for the assumption that the Fast and Accurate score were capable of explaining the Efficient score. The reason for this relationship being absent remains an open question.