Journal of the Association for Information Systems

JAIS

Special Issue Editorial

Towards a NeuroIS Research Methodology: Intensifying the Discussion on Methods, Tools, and Measurement

René Riedl
University of Applied Sciences Upper Austria and University of Linz
rene.riedl@fh-steyr.at

Fred D. Davis
University of Arkansas
fdavis@walton.uark.edu

Alan R. Hevner
University of South Florida
ahevner@usf.edu

Abstract

The genesis of the Neuro-Information Systems (NeuroIS) field took place in 2007. Since then, a considerable number of IS scholars and academics from related disciplines have started to use theories, methods, and tools from neuroscience and psychophysiology to better understand human cognition, emotion, and behavior in IS contexts, and to develop neuro-adaptive information systems (i.e., systems that recognize the physiological state of the user and that adapt, based on that information, in real-time). However, because the NeuroIS field is still in a nascent stage, IS scholars need to become familiar with the methods, tools, and measurements that are used in neuroscience and psychophysiology. Against the background of the increased importance of methodological discussions in the NeuroIS field, the Journal of the Association for Information Systems published a special issue call for papers entitled "Methods, tools, and measurement in NeuroIS research" in 2012. We, the special issue's guest editors, accepted three papers after a stringent review process, which appear in this special issue. In addition to these three papers, we hope to intensify the discussion on NeuroIS research methodology, and to this end we present the current paper. Importantly, our observations during the review process (particularly with respect to methodology) and our own reading of the literature and the scientific discourse during conferences served as input for this paper. Specifically, we argue that six factors, among others that will become evident in future discussions, are critical for a rigorous NeuroIS research methodology; namely, reliability, validity, sensitivity, diagnosticity, objectivity, and intrusiveness of a measurement instrument. NeuroIS researchers—independent from whether their role is editor, reviewer, or author—should carefully give thought to these factors. We hope that the discussion in this paper instigates future contributions to a growing understanding towards a NeuroIS research methodology.

Keywords: *Brain, Diagnosticity, Intrusiveness, Methodology, Methods, NeuroIS, Neuroscience, Nervous System, Objectivity, Psychophysiology, Reliability, Sensitivity, Validity.*

Volume 15, Special Issue, pp. i-xxxv, October 2014

Towards a NeuroIS Research Methodology: Intensifying the Discussion on Methods, Tools, and Measurement

1. Introduction

During the past decade, increasingly more scholars from the social and economic sciences and from computer science have started to use methods and tools from Neuroscience. This development is expected to result in a better theoretical understanding of human behavior such as decision making. Moreover, using Neuroscience methods and tools may contribute to the design and development of innovative information systems as demonstrated, for example, by neuro-adaptive information systems (e.g., Astor, Adam, Jerčić, Schaaff, & Weinhardt, 2014) and affective computing applications (see, e.g., the papers published in *IEEE Transactions on Affective Computing,* a journal launched in 2010). Against the background of the increased use of Neuroscience methods and tools in scientific fields that are closely related to the Information Systems (IS) field, scholars have introduced the concept of NeuroIS into the IS literature (Dimoka, Pavlou, & Davis, 2007; Riedl, 2009). In essence, NeuroIS is a subfield in the IS field that uses neuroscience and neurophysiological methods, tools, and theories to better understand the design, development, and use of information and communication technologies (ICT) in organizations and society. Specifically, NeuroIS is expected to contribute to the development of new theories that make possible accurate predictions of ICT-related behavior, and to the design of ICT artifacts that positively impact economic and non-economic variables such as productivity, satisfaction, adoption, and well-being (Riedl et al., 2010a, p. 245).

Because the NeuroIS field is still in a nascent stage (even though several empirical contributions such as Dimoka (2010), Léger, Riedl, and vom Brocke (2014), Ortiz de Guinea, Titah, and Léger (2014), Riedl, Hubert, and Kenning (2010b), and Riedl, Mohr, Kenning, Davis, and Heekeren (2014) already exist), IS scholars need to become familiar with the methods, tools, and measurements that are used in Cognitive Neuroscience and in related disciplines (e.g., Psychophysiology). Based on a higher degree of familiarity, IS academics (editors, reviewers, and authors) can develop sound methodological knowledge that is necessary to evaluate whether or not a specific method, tool, or measurement is suitable to study a specific IS research question and whether a method or tool is correctly applied. Without such a knowledge base, IS scholars cannot leverage the full potential of Neuroscience for IS research because the production of scientific knowledge depends to a great extent on the techniques for collecting, analyzing, and interpreting data and the ways in which the techniques are applied (Simon, 1980).

To date, scholars have identified several methods, tools, and measurements as useful for investigating IS research questions. Dimoka et al. (2012) and Riedl et al. (2010a), for example, comprehensively overview current NeuroIS methods, tools, and measurements that range from physiological methods (e.g., pupillometry, electrocardiogram (EKG), facial electromyography (fEMG), or electrodermal activity (EDA) measurement) to tools that measure brain activity or related physiological activity (e.g., functional magnetic resonance imaging (fMRI), electroencephalography (EEG), or near infrared spectroscopy (NIRS)). These papers also describe techniques that are based on brain morphology (e.g., lesion studies, voxel-based morphometry (VBM), or diffusion tensor imaging (DTI)). Moreover, recent experiments illustrate the potential of measuring hormones and related substances such as enzymes for IS research (Riedl, Kindermann, Auinger, & Javor, 2012; Riedl, 2013; Tams, Hill, Ortiz de Guinea, Thatcher, & Grover, 2014). In addition to method overviews, fMRI guidelines (Dimoka, 2012) and more general guidelines for Neuroscience studies in IS research (vom Brocke & Liang, 2014) have been recently published in mainstream IS journals. Accordingly, methodological contributions and discussions have already started to emerge in the NeuroIS literature, which has contributed to an increased understanding of, and interest into, a NeuroIS methodology. However, despite these first papers, more research contributing to the systematic development of a NeuroIS research methodology is needed.

In the past, IS researchers have often relied on survey and interview data. While these techniques have contributed to theoretical advancements, Neuroscience methods, tools, and measurements are expected to be less biased because self-reported data are susceptible to, among others, common method, social desirability, and subjectivity biases (e.g., Dimoka et al., 2011, p. 688; vom Brocke et al., 2013). Moreover, computer users' feelings often do not reach the level of awareness, and therefore it is not possible to report on them in survey or interview studies. Consequently, Neuroscience offers great potential to investigate feelings and similar user states during human-machine interaction and in other IS contexts (see, for example, Dimoka et al., 2012), which can complement traditional approaches.

Reflecting the potential that Neuroscience methods, tools, and measurements offer IS research, the *Journal of the Association for Information Systems* published a special issue call for papers entitled "Methods, tools, and measurement in NeuroIS research" in 2012 in order to foster methodological contributions to the NeuroIS literature such as philosophical considerations and measurement issues to more specific aspects of data collection and analysis. Considering the importance of both theoretical research and design science research in the IS field, along with a research culture that does not dismiss a paper simply due to a specific methodological focus, the call for papers explicitly indicated that contributions related to all relevant methods, tools, and measurements and to research that is located at the nexus of Neuroscience and both behavioral research and design science research would be welcome.

Altogether, we received twenty papers of which we accepted three. Table 1 summarizes the accepted papers. A special issue advisory and editorial board consisting of both IS scholars (Henri Barki, Samir Chatterjee, Alan Dennis, David Gefen, Pierre-Majorique Léger, Adriane Randolph, Eric Walden) and researchers from other fields (Peter Kenning: Consumer Neuroscience and Marketing, Gernot Müller-Putz: Brain-Computer Interfacing and Computer Science, Martin Reuter: Biological Psychology and Behavioral Genetics) helped to evaluate the manuscripts' quality. Specifically, members of this board served as associate editors for the manuscripts in their respective field of expertise and provided support to find high-quality reviewers. To create a rich picture of the potentially diverse perceptions on the quality of manuscripts, we attached importance to "interdisciplinary review teams", and submissions were typically reviewed by one or two IS scholars and one or two non-IS scholars whose scientific background included, among others, Cognitive Neuroscience, Biology, and Statistics.

Against the background of our observations during the review process (in particular regarding the methodological issues that the reviewers and associate editors raised), our own reading of the literature relevant to methodology, and the statements and comments provided during a panel discussion on "Key criteria for NeuroIS research"[1], we decided to share some of our current thoughts on a NeuroIS research methodology. We hope that the discussion in this paper will direct readers' attention to important methodological aspects in NeuroIS research and thereby 1) sensitize journal editors, reviewers, and authors to themes that have significant influence on methodological quality and 2) affect the overall quality of research findings and the conclusions drawn from NeuroIS research.

[1] The panel discussion took place on June 7 during the Gmunden Retreat on NeuroIS 2014 (www.neurois.org).

Table 1. Summary of Papers in this Special Issue	
Paper	**Major contribution**
Léger, P.-M., Sénecal, S., Courtemanche, F., Ortiz de Guinea, A., Titah, R., Fredette, M., Labonte-Lemoine, È. (2014). Precision is in the eye of the beholder: Application of eye fixation-related potentials to information systems research.	This paper introduces the eye-fixation related potential (EFRP) method to IS research. This method allows researchers to synchronize eye tracking with EEG recording to precisely capture users' neural activity at the time at which they start to cognitively process a stimulus (e.g., event on the screen in a human-computer interaction task). This complements and overcomes shortcomings of the traditional event related potential (ERP) method, which can only stamp the time at which a stimulus is presented to a user. The authors illustrate the EFRP method with an experiment in a natural IS use context.
Vance, A., Anderson, B., Kirwan, C. B., & Eargle, D. W. (2014). Using measures of risk perception to predict information security behavior: Insights from electroencephalography (EEG).	Previous research on perceptions of information security risk has primarily relied on self-reports. However, because risk perceptions are often associated with feelings (e.g., fear or doubt) that are difficult to measure accurately using survey instruments, this paper contributes by demonstrating that risk-taking behavior is effectively predicted using EEG via event-related potentials (ERPs). The paper compares the predictive validity of EEG measures to that of self-reported measures of information security risk perceptions.
Tams, S., Hill, K., Ortiz de Guinea, A., Thatcher, J., & Grover, V. (2014). NeuroIS—alternative or complement to existing methods? Illustrating the holistic effects of Neuroscience and self-reported data in the context of technostress research.	While some previous studies indicate that NeuroIS constitutes an alternative to self-reports, which implies that the two methods assess the same dimension of an underlying IS construct, other studies indicate that NeuroIS constitutes a complement to self-reports, which implies that the two methods assess different dimensions of an IS construct. To clarify this issue, this paper examines whether NeuroIS and psychometrics constitute alternatives or complements. The authors conduct their examination in the context of technostress.

This paper is structured as follows. In Section 2, we provide fundamentals on human neurobiology because at least a basic understanding about the human nervous system is necessary to comprehend methods, tools, and measurement in NeuroIS research. In Section 3, we define the concept of methodology and, in Section 4, we present a framework that researchers can use to structure a methodology discussion in NeuroIS research. Subsequently, in Section 5, we discuss selected aspects of a NeuroIS research methodology that we consider of particular importance in the current stage of the field. Specifically, we discuss six factors: reliability, validity, sensitivity, diagnosticity, objectivity, and intrusiveness. Finally, in Section 6, we conclude the paper.

2. Fundamentals of Human Neurobiology

In this section, we briefly introduce human neurobiology. Specifically, we summarize fundamentals on human physiology at a high level of abstraction. This introduction provides a basis to better understand the sections to follow.

Without a nervous system, it would neither be possible for humans to perceive the external environment, nor could feelings, thoughts, and behaviors emerge. The human nervous system consists of different parts. At a high abstraction level, we can distinguish the central nervous system (CNS) and the peripheral nervous system (PNS); the former is sub-divided into the brain and spinal cord, and the latter comprises all neural tissue except for the CNS. The PNS can be further sub-divided into the somatic nervous system (SNS) and the autonomic nervous system (ANS). The ANS, in turn, consists of the sympathetic division (activates the body) and parasympathetic division (relaxes the body). Importantly, even though different parts of the nervous system can be separated anatomically (as Figure 1 illustrates), their functions are highly interrelated. Thus, for humans to be healthy and perform well, successful interaction among the different parts of the nervous system is critical (along with their proper interaction with other tissue in the human body (e.g., organs)). It has been argued that the brain (i.e., the information processing unit) and the ANS (i.e., the unit that keeps the body in balance; this balance is referred to as homeostasis) are the major units of analysis in NeuroIS research, while the spinal cord and the SNS are less important (see, for example, a paper on technostress by Riedl (2013) for details regarding these differences in relative importance).

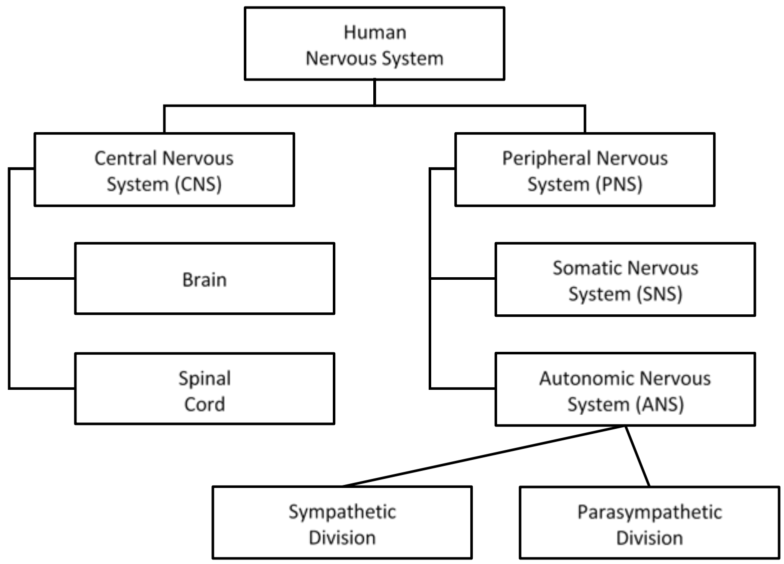

Figure 1. Major Parts of the Human Nervous System

The parts of the human nervous system serve different functions. The CNS's major functions include integrating, processing, and coordinating sensory information, and coordinating motor commands. Different types of receptors exist in the human body, and their purposes are to collect information and transduce this information to bioelectric signals, which are subsequently transmitted to the brain. Transmissions may either occur via the spinal cord (e.g., receptors in the skin provide information on pressure or temperature) or directly in the head (e.g., specialized receptors for the provision of visual information based on light perception). Depending on the type of information and other factors (e.g., context), different brain regions perform integration and subsequent information processing. What follows is that the brain sends commands to peripheral systems (e.g., extremities or organs) either via the PNS (specifically the motor part) or via hormones and precursor substances of hormones (i.e., chemicals released by glands or cells that act as messengers in the body). Depending on the type of effectors, the ANS and/or the SNS become active; the former primarily regulates heart rate, smooth muscles (e.g., to control eye movements), and glands (e.g., to release behaviorally relevant hormones); the latter controls skeletal muscles. Generally, components of the nervous system travelling from peripheral systems to the brain are referred to as afferents, and components travelling from the brain to peripheral systems are referred to as efferents.

The core element of the nervous system is the neuron, which is an electrically excitable nerve cell that receives, processes, and sends information. Operation of a neuron is based on electrical impulses, and communications with other neurons occur through chemical signals. A typical neuron consists of a cell body (referred to as soma), dendrites, and an axon. Information processing takes place in the soma, receipt of information occurs via dendrites, and an axon sends information. Typically, a cell body has multiple dendrites, but only one axon.

When stimulation of a nerve cell reaches a threshold, an electrical impulse is transmitted to another neuron. Specifically, the electrical signal passes along the axon, which causes the release of neurotransmitters from terminal buttons into a synapse. This phenomenon is referred to as action potential or neural firing. A fundamental principle underlying neural firing is that it works on an all-ornone-basis, which means that a neuron cannot partially fire. What follows is that a neuron either fires or does not fire, and the firing of a neuron cannot be strong or weak. How often a neuron fires is a function of stimulation strength, which means that it is dependent on the number and frequency of excitatory and inhibitory signals[2]. The source underlying operation of an afferent neuron may be another neuron or a receptor, and the target of an efferent neuron may be another neuron, a skeletal muscle, or an internal organ.

3. What Does Methodology Mean?

In this paper, we provide our thoughts on a NeuroIS research methodology. But what is the exact meaning of the term "methodology"? The "Merriam-Webster Encyclopedia" indicates two different, yet related, meanings. First, "a body of methods, rules, and postulates employed by a discipline: a particular procedure or set of procedures". Second, "the analysis of the principles or procedures of inquiry in a particular field". Thus, methodology deals with the methods, tools, and measurement in a scientific field, and also with systematic examination of the employed methodological practices (i.e., meta-research). Importantly, a methodology's utility in a field cannot be assessed without considering the field's subject matter (in the IS field: socio-technical systems in organizations and society) and the goals of scientific inquiry (i.e., primarily explanation and design). Thus, a methodology that is appropriate in one field might not be in another one.

Consistent with the above stated meanings, in this paper, we adopt a definition by Mingers (2001) who accurately conceptualizes methodology as "a structured set of guidelines or activities to assist in generating valid and reliable research results" (p. 242). The major purpose for a scientific field to have a methodology is to establish common understanding among editors, reviewers, and authors on "what" constitutes good and poor, or acceptable and unacceptable, methodological quality, and "why". Moreover, guidelines are needed for documenting key criteria and normative standards for NeuroIS research in general and for specific measurement instruments[3].

In this context, it is important to reference a seminal paper on normative standards for IS research (Straub, Ang, & Evaristo, 1994). What is most interesting from a methodological perspective is that a factor analysis revealed that, out of fifteen "key criteria for high-quality research", four criteria (replication, statistical/mathematical analysis, research design, and scientific ethics) loaded on one factor, referred to as "conduct of research". It was exactly this factor that explained the highest proportion (26.1%) of manuscript quality perceptions among a sample of 144 IS professionals (i.e., authors and editorial board members). The sample rated other factors (presentation of research: 20.4%, conceptual significance: 16.9%, practical significance: 14.9%, and reputation of the author and his or her institution: 12.3%) less importantly. Straub et al. (1994) define "conduct of research" as follows (p. 34):

> *Methods, subjects, and techniques are well suited to the exploration of the research questions. The work demonstrates appropriate operationalizations of theoretical constructs and an acceptable degree of internal and/or external validity. The choice of statistical and/or mathematical analysis is appropriate, as is the interpretation of results. Study results are objective and in such a form that other researchers could replicate the work. The work adheres to generally accepted standards for scientific ethics.*

We draw six conclusions from the Straub et al. (1994) paper that are directly relevant for the present paper. First, methodological quality in general is the most important factor for assessing manuscript quality (a fact that has most likely not changed during the past two decades, at least not for top journals). Second, there must be a good fit between the research question and methods. Third, a clear and unquestionable link between theory and measurement of the real-world phenomenon (i.e.,

---

[2] Signals that arrive at dendrites can be either excitatory or inhibitory; while excitatory signals depolarize the cell membrane and inhibitory signals hyperpolarize the cell membrane (Gazzaniga, Russell, & Senior, 2010). Depolarization increases a neuron's firing probability, while hyperpolarization decreases firing probability.

[3] Different guidelines for conducting and evaluating different types of research have been published in the IS literature. Venkatesh, Brown, and Bala (2013), in their guideline paper on mixed methods research, list several other IS guideline papers (see p. 22). Moreover, as already mentioned, guideline papers also are emerging in the NeuroIS field; see Dimoka (2012) and vom Brocke & Liang (2014).

operationalization) is essential. Fourth, type of data affects data analysis, a fact particularly important in NeuroIS research. Fifth, replication should be possible, a fact that is primarily relevant in quantitative research and hence also in the NeuroIS field. Sixth, adherence to scientific ethics is indispensable, a fact which is also highly relevant in NeuroIS research. These conclusions substantiate the need for a clear and comprehensive NeuroIS research methodology.

4. Guiding Framework for the Present Paper

Based on our conceptualization of methodology and our brief reflection on the results and implications of Straub et al.'s (1994) study on normative standards in IS research, in this section, we present a framework that we use to structure our discussion of methodology in NeuroIS research. This framework is based on the notion that NeuroIS research holds significant potential for theoretical research that aims to understand and explain phenomena in reality; note that this does not rule out that NeuroIS also holds potential for design and engineering initiatives (see, e.g., Loos et al., 2010; vom Brocke, Riedl, & Léger, 2013; Hevner, Davis, Collins, & Gill, 2014). Figure 2 illustrates the framework. We explicitly note that the framework rests on the assumption that the research includes an empirical part.

In essence, the framework summarizes basic research activities. Theoretical research as defined here refers to some phenomenon of interest, which defines the research question. For example, a researcher could be interested in the nature of technostress and its antecedents and consequences. Once the research question has been fixed, a literature review and theorizing process follows. Next, the research reaches a level where methodology comes into primary focus (see the part with the gray background color in Figure 2). Specifically, the theoretical model and the constructs have to be transformed into a research design and measurable variables, and this also includes the exact planning of the methods to be employed. What follows are data collection and analysis. An interpretation of the research findings comes next, ultimately resulting, at least in most research projects, in the publication of the study. The contribution is typically a new or revised theory (i.e., changed knowledge about reality), and this affects the phenomena to be studied in future research.

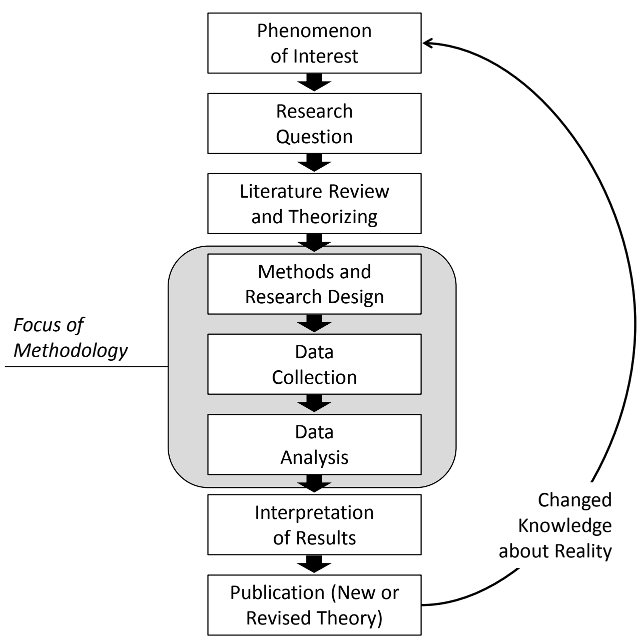

Figure 2. Framework to Structure the Discussion on Methodology in NeuroIS Research

As Figure 2 shows, theoretical research follows a process of activities. In practice, however, the activities may not necessarily follow a strictly sequential process, but may also include iterative loops. The framework also shows the focus of a discussion on methodology; namely, methods and research design, data collection, and data analysis, and that methodology is closely related to theorizing (input) and interpretation of results (output).

5. Selected Aspects of a NeuroIS Research Methodology

In this section, we discuss selected key aspects of a methodology of NeuroIS research. Despite the fact that we consider the following methodological aspects to be important, we note that future papers on a NeuroIS methodology will identify further aspects and thus broaden the scope of the present discussion. Moreover, we hope that future papers will delve into methodological details related to the aspects that we discuss here and thus deepen our insights. Consequently, we consider the present paper as a starting point towards an emerging understanding of a rigorous NeuroIS research methodology. Importantly, as can be inferred from the discussions to follow in this paper, the complexity of all presented methodological aspects is high, which implies that comprehensive and complete examination of these aspects would fill entire textbooks and journals. Thus, this paper creates awareness for important methodological topics and questions to intensify the recently started discourse.

Information Systems is an applied scientific field. Consequently, IS researchers are primarily interested in real-world phenomena. Reality can have three different forms; see, for example, an essay by Mingers (2001), based on Habermas (1990). First, the "material world" exists independently of humans' subjective experience, and is therefore objective. It existed before the emergence of humans and most likely will exist after the human race's extinction. The material world is the research object in natural sciences (e.g., Physics, Chemistry, Biology). Second, the "personal world" comprises beliefs, attitudes, thoughts, and experiences, among other phenomena. This world is subjective and cannot exist without the individual who experiences it. Third, a "social world" also exists. Humans are typically in interaction with other humans, and they participate in social life. The social world is intersubjective because it is a subjective mental construction on the one hand, but one that goes beyond the individual on the other hand. The personal world and the social world are the research objects in social sciences (e.g., sociology).

Information Systems research is concerned with socio-technical systems, and hence the field is primarily a social science. However, this fact does not rule out the possibility of influences from theoretical sciences such as computer science, mathematics, or logic (Vogel & Wetherbe, 1984), among other influences. Moreover, the IS field is, as demonstrated by the recent genesis and development of the NeuroIS field, increasingly influenced by theories, methods, tools, and measurement from sciences with a focus on the material world such as Neuroscience (e.g., Dimoka, Pavlou, & Davis, 2011; Dimoka et al., 2012; Riedl et al., 2010a), Genetics, Neurobiology, and Endocrinology (e.g., Riedl et al., 2012; Riedl, 2013). Thus, operationalization of constructs in IS research may happen on different levels, ranging from physiology (i.e., the material world) to beliefs, attitudes, behavioral intentions, and actual behavior (i.e., the personal and social worlds).

Figure 3 summarizes major methodological aspects that have to be considered in NeuroIS research. Generally, the real-world phenomena related to IS research are conceptualized as more or less abstract mental representations, referred to as constructs in theoretical models. Because theoretical research seeks to establish causalities among constructs, theoretical models usually also indicate causal directions of relationships among constructs. Constructs, "a conceptual term used to describe a phenomenon of theoretical interest" (Edwards & Bagozzi, 2000, p. 155), typically cannot be observed directly. Thus, scientific investigation of a construct implies operationalization, and measurement may refer to different analytical levels (e.g., physiology, belief, attitude, behavioral intention, or behavior). For the purpose of measurement, instrumentation is needed. Specifically, depending on the measure, instrumentation may range from tools to capture physiological activity in the brain (e.g., fMRI) and survey instruments (to capture beliefs, attitudes, or behavioral intentions), to tools that allow for collection of data on actual behavior (e.g., video camera, or software to capture clickstreams and mouse navigation patterns). Thus, through application of instrumentation, a researcher can capture measures, defined as "observed score[s of variables] gathered through self-report, interview, observation, or some other means" (Edwards & Bagozzi, 2000, p. 156)[4]. Finally,

[4] Note that a variable is defined as "an observable entitiy which is capable of assuming two or more values" (Bacharach, 1989, p. 500).

Figure 3 indicates that the relationship between construct and measure can be either reflective or formative. If variation in a construct results in variation in its measures, then measures are called reflective; in contrast, if measures are viewed as causes of constructs, then measures are called formative (for detailed discussions, see, for example, Edwards and Bagozzi (2000) or Bagozzi (2011)).

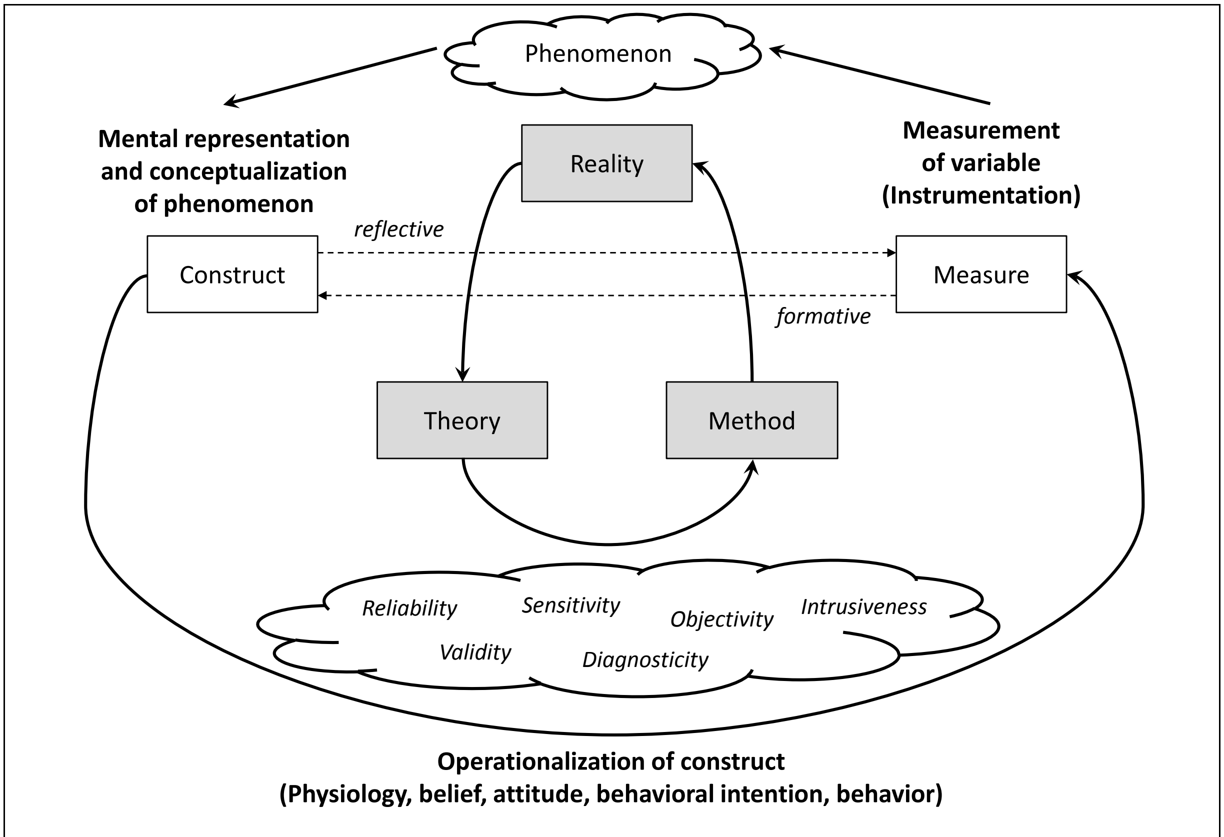

Figure 3. Major Methodological Aspects in NeuroIS Research with Theory Focus

Considering the basic functioning of theoretical research (as illustrated in Figure 2), several methodological themes emerge, six of which (reliability, validity, sensitivity, diagnosticity, objectivity, and intrusiveness) we discuss in the following sections, based on illustrative IS research examples. Before we do so, we provide some notes on the conceptualization of a phenomenon, a major task in empirical research, and hence also in NeuroIS studies, when a real-world phenomenon has to be represented mentally in the researcher's mind; that is, when it has to be transferred from the real to the theoretical world.

To conceptualize means "to form an idea, picture, etc. of something in [the] mind" ("Merriam-Webster Encyclopedia"). A major challenge of empirical research is the development of adequate conceptualizations of real-world phenomena. Clark and Watson (1995, p. 310) write:

> *A critical first step is to develop a precise and detailed conception of the target construct and its theoretical context. We have found that writing out a brief, formal description of the construct is very useful in crystallizing one's conceptual model.*

An adequate definition of the conceptual domain of a construct is essential for at least three reasons (MacKenzie, Podsakoff, & Podsakoff, 2011, p. 295). First, a good construct definition helps to clarify what the construct does and does not refer to, and this is a precondition to make similarities and differences between the focal construct and other constructs explicit. Second, an adequate definition is important because it reduces the probability that a construct's indicators are inadequate or contaminated because the construct overlaps with other constructs. Third, a good definition reduces the probability of invalid conclusions about relationships among constructs.

Many constructs in IS research should be conceptualized, and consequently operationalized, not as purely cognitive, behavioral, or biological phenomena, but rather as a blend. Consider the example of technostress (for a review, see Riedl, 2013). This construct has been defined in seminal publications as “a modern disease of adaptation caused by an inability to cope with the new computer technologies in a healthy manner” (Brod, 1984, p. 16), or as “any negative impact on attitudes, thoughts, behaviors, or body physiology that is caused either directly or indirectly by technology” (Weil & Rosen, 1997, p. 5). Clearly, an IS researcher interested in investigating technostress should consider these existing definitions and decide whether they adequately represent the phenomenon and construct of interest. Moreover, existing definitions from the literature could be complemented by further specifications that result from interviews with practitioners and experts.

From the perspective of a NeuroIS methodology, the critical aspect here is that many IS phenomena (e.g., technostress) are so broad that their conceptualization necessarily has to consider multiple analytical levels in order to fully capture the phenomenon. While such a call for multilevel conceptualization and operationalization of IS constructs is definitely reasonable from a NeuroIS viewpoint, the concrete realization in IS research practice might not be so easy. One major reason for this conclusion is that the existing research culture in the IS field is one that has not extensively focused on multilevel conceptualizations. In a paper on construct measurement, MacKenzie et al. (2011) write:

> [T]*he definition should specify whether the construct refers to a thought (e.g., cognition …), a feeling (e.g., … emotion …), a perception (e.g., perceived ease of use of technology …), an action (e.g., behavior …), an outcome (e.g., degree of use …), or an intrinsic characteristic (e.g., cognitive ability …)* (p. 298).

We absolutely agree with MacKenzie et al. (2011) that careful construct specification is critical, and add that future IS research should consider complementary physiological conceptualization of constructs whenever reasonable.

In the following sections, we discuss, based on illustrative IS research examples, the critical concepts of reliability, validity, sensitivity, diagnosticity, objectivity, and intrusiveness from a NeuroIS perspective.

5.1. Reliability

Reliability refers to the proportion of a measurement’s total variance that is due to random error, rather than true variance of the underlying construct. Unreliability attenuates observed correlations, path coefficients, and tests of differences. Reliability indicates the extent to which a measurement instrument is free of measurement error, and therefore yields the same results on repeated measurement of the same construct. In fact, test-retest correlations are commonly used to assess reliability (as are other methods such as internal consistency of multiple measures of a construct). In IS research, many scholars have made reliability an important subject of discussion, primarily because measurement error, if too high, discredits research findings. Since a large proportion of construct measurement is based on survey instruments in the IS field, scholars have discussed reliability issues mostly in the context of survey research. If respondents are inconsistent in their (repeated) answers to the items reflecting a theoretical construct, measurement error is high and precision is low, and consequently measurement is unreliable. Generally, in survey research, the investigator attempts to find “proximal measures of the true scores” that describe the phenomenon (Straub, Gefen, & Boudreau, 2004) because it is unlikely, or impossible, that the true score can be found. Several publications in the IS field (e.g., Straub, 1989; Straub et al., 2004) have described techniques to assess a survey instrument’s reliability (e.g., test-retest, internal consistency), and application of these techniques is standard in mainstream IS research today.

In the context of NeuroIS research, the fundamental question of measurement reliability also has to be raised: are neurobiological measures stable across repeated measurements? To the best of our knowledge, this important question on the reliability of neurobiological measurement has not been explicitly addressed in IS research so far. It is possible that IS researchers assume that neurobiological measurement is reliable per se, or at least generally more reliable than survey data. Coming back to the question whether neurobiological measures are (necessarily) stable across repeated measurement and hence (perfectly) reliable, the answer is that they are not. According to our observations, this fact is neither well known in the IS field, nor in several other fields that use the prefix “neuro“.

In Section 2, we indicate that messages in the human body are often transmitted via hormones. Moreover, it has been shown that specific hormones, such as those related to stress (e.g., cortisol), might be important measures in IS research (Riedl, 2013). It is tempting to assume that physiological measurement is reliable per se, or generally more reliable than other measures. For example, one might conclude that hormone assessments are perfectly or highly reliable. However, they are not. In endocrinological research, the concept of precision is closely related to measurement reliability. Precision is “the closeness of agreement between test results repeatedly and independently obtained under stable conditions” (Schultheiss & Stanton, 2009, p. 27). Table 2, based on Schultheiss and Stanton, summarizes standard procedures used in endocrinological research to calculate precision of hormone measurements and provides an example. Generally, NeuroIS scholars using endocrinological measurement should consider precision in their research. We also note that outsourcing hormone analysis to a specialized laboratory may affect the possibility to establish precision and accuracy (discussed in Section 5.2.). Further details are available in Schultheiss and Stanton (2009) and in the references they cite.

Table 2. Precision in Hormone Research	
Definition	The closeness of agreement between test results repeatedly and independently obtained under stable conditions.
Procedure	Variation across multiple measurements, calculated based on the mean (M) of multiple measurements and the standard deviation (SD) of the measurements.
Metric	Precision = M/SD×100. Intra- and inter-assays coefficients of variation of less than 10% are considered good.
Example	Cortisol: Intra-assay coefficient: 3.42%, Inter-assay coefficient: 6.90%, Source: Riedl et al. (2012, p. 66).

For measuring hormones, several further issues are critical, whose detailed discussion is beyond the scope of the present paper. However, the following examples are indicative of the methodological knowledge needed to plan and execute endocrinological studies in a reliable way.

Can a specific hormone be measured reliably in saliva, or is it necessary to draw blood samples? Obviously, taking blood samples might significantly influence participants‘ physiological, emotional, and cognitive states (e.g., by increasing stress) and thereby potentially bias the measurement of a variable. With respect to measuring the stress hormone cortisol, it is a well-established fact that cortisol levels measured in saliva are similar to free cortisol levels measured in blood and cortisol levels in the brain (e.g., $r > 0.90$ between saliva and blood, Foley & Kirschbaum, 2010). Thus, taking saliva samples of cortisol can be considered reliable. However, there is an ongoing debate whether several other behaviorally relevant hormones (which are related to important IS constructs) can be measured reliably in saliva. As an example, oxytocin, a neuroactive substance related to trust (for a review, see Riedl & Javor, 2012), cannot be measured reliably in saliva (e.g., Carter et al., 2007; Horvat-Gordon, Granger, Schwartz, Nelson, & Kivlighan, 2005; McCullough, Churchland, & Mendez, 2013).

The issue of reliability also concerns other physiological measures, including those related to both the CNS (e.g., fMRI, EEG, TCD) and ANS (e.g., heart rate, blood pressure, skin conductance). Suppose you are a NeuroIS scholar and you wish to study users‘ activation and stress that result from perception of computer hassles (e.g., system breakdown or long response time). One major physiological indicator would be blood pressure (BP). This indicator measures the pressure exerted by circulating blood on the walls of blood vessels. During each heartbeat, BP varies between a minimum (diastolic) and maximum (systolic). A person‘s BP is usually expressed in terms of the systolic over diastolic pressure, and is measured in millimeters of mercury (mmHg). A value of 120/80, for example, is typical for a healthy adult. A technique often used to measure blood pressure is called the auscultatory method, which uses a stethoscope and a sphygmomanometer (see Figure 4).

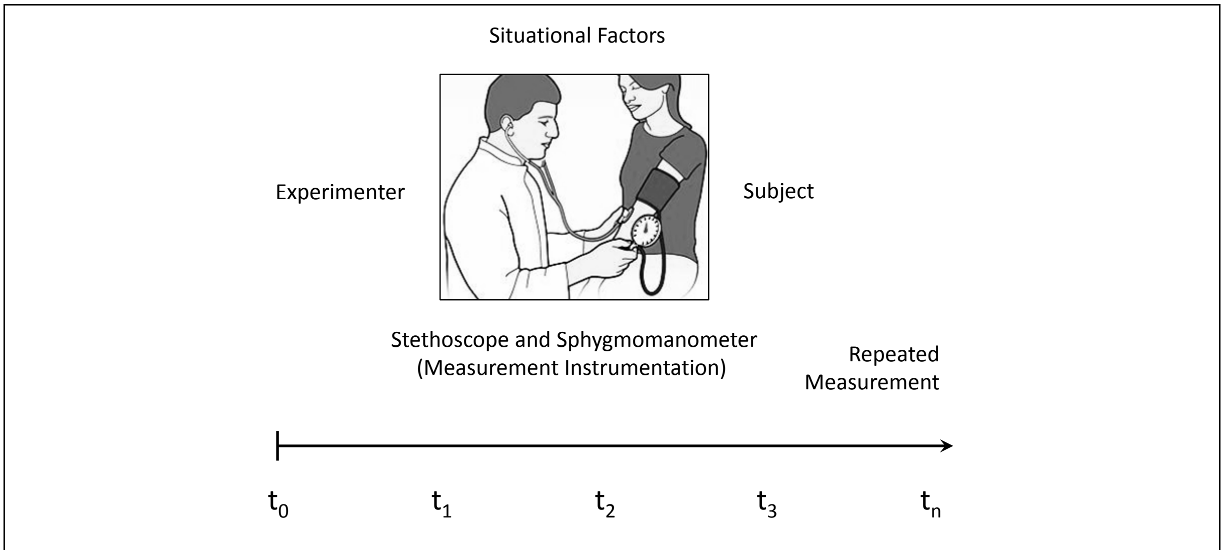

Figure 4. Sources of Measurement Error: The Example of Blood Pressure Measurement

A researcher using BP as an indicator of activation and stress that results from human-computer interaction could be interested in the reliability of both diastolic and systolic BP. The ability to detect changes in BP relies on knowing a subject's true BP, but that true value might be obscured by several other factors (Strube & Newman, 2007). Figure 4 conceptually illustrates major influencing factors, including those related to (i) measurement instrumentation, (ii) the experimenter (e.g., whether they are trained to use the instrumentation, or whether their perception functions properly, a fact that is relevant in the example because acoustic and optic stimuli must be processed), (iii) situational factors (e.g., a subject having too much coffee before the experiment), and (iv) subject-related factors (e.g., health status, age, or sex). Because a researcher typically wants to study variability in BP as a function of stimulus perception (here computer hassles), factors creating variability in BP unrelated to the stimulus are potential sources of measurement error that will possibly reduce reliability.

Test-retest reliability is a well-known technique to assess the reliability of measurement instrumentation. This technique is not only used to validate survey instruments, but also holds significant value in determining the reliability of physiological measures (some example studies are discussed below). This technique calculates the correlation between the same measure on at least two different occasions, and hence it is based on the logic of repeated measurement at different times (t_0 through t_n, Figure 4). Generally, it is assumed that measurements taken from the same subject under the same conditions at two different times (e.g., BP at t_1 and at t_2) differ only in their random components so that the correlation between the scores at t_1 and at t_2 reveals the relative amounts of true score variance that the measures contain (Straub et al., 2004; Strube & Newman, 2007). Importantly, this logic rests on the assumption that the true BP score remains stable between t_1 and t_2. If this assumption is not realistic in the context of a specific study, then the test-retest reliability estimate should not be used as an indicator of measurement error.

Altogether, Figure 4 shows, based on the example of BP recording, that measurement of physiological variables might be unreliable due to different sources of measurement error, including factors related to the measurement instrument, experimenter, situational factors, and the subject. Moreover, the timeline at the bottom signifies that test-retest measurement is a standard procedure in physiology to establish reliability. Of course, depending on the instrumentation used to determine a subject's physiological state, sources of measurement error may vary. For example, unlike studies using BP measurement based on the auscultatory method, in investigations based on other BP measurement methods (e.g., digital, oscillometric monitors), or in studies measuring other physiological signals such as the blood-oxygen-level dependent (BOLD) signal (underlying fMRI), it is less possible that the experimenter directly constitutes a source of measurement error because signal recording is purely machine-based and hence independent from perception and/or interpretation of the experimenter[5]. Thus, a NeuroIS researcher should carefully consider and manage the potential sources of measurement error in the context of a specific study to increase measurement reliability.

With respect to reliability of the BOLD signal, several studies have been carried out in Neuroscience, Biology, and Medicine. Regarding the reliability of task-dependent brain signals, evidence shows that test-retest reliability of evoked BOLD signals is basically good (Plichta et al., 2012). Specifically, Plichta et al. examined within-subject and group-level reliability of different tasks (emotional, motivational, and cognitive). A total of 25 healthy subjects were scanned twice on a 3T MRI scanner with a mean test-retest interval of 14.6 days. The results indicate that robust activation of all three tasks was found in expected target regions (emotional task: amygdala, motivational task: striatum, and cognitive task: prefrontal cortex). Moreover, it is reported that "[r]eliability of group level activation was excellent for all three tasks with ICCs [intraclass correlation coefficients[6]] of 0.89-0.98 at the whole brain level and 0.66-0.97 within target ROIs [regions of interest]" (p. 1746). Within-subject reliability of ROI-mean amplitudes across sessions was fair-to-good for the motivational (monetary reward anticipation) and cognitive (n-back working memory) tasks, but lower for the emotional task (face matching). Reflecting on their results, the authors conclude that "fMRI reliability characteristics can be strongly dependent on the nature of the fMRI task", and that their study "provides task-specific fMRI reliability performance measures that will inform the optimal use, powering and design of fMRI studies using comparable tasks" (p. 1746).

Regarding reward processing, Fliessbach et al. (2010) conducted an fMRI experiment to examine test-retest reliabilities of BOLD responses to reward prediction, reward receipt, and reward prediction errors (with a focus on the ventral striatum and the orbitofrontal cortex). They investigated a total of 25 subjects based on different reward-related tasks with a test-retest interval of 7-13 days. The results indicate that, on a group level, the task resulted in significant activations of the respective brain areas in the two sessions. However, test-retest reliabilities were poor to fair (Fliessbach et al. report ICCs of -0.15 to 0.44). Moreover, they reported that ICCs for motor activations were considerably higher; the range of ICCs was 0.32 to 0.73. They conclude that "results reveal the methodological difficulties behind across-subject correlations in fMRI research on reward processing. These results demonstrate the need for studies that address methods to optimize the retest reliability of fMRI" (p. 1168).

Another important issue in fMRI research is the reliability of multisite BOLD data (i.e., data captured in different machines). In this context, Brown et al. (2011) argue that researchers perform multi-site functional MRI examinations to enhance statistical power and generalizability; however, they also indicate that "undesired site variation in imaging methods could off-set these potential advantages" (p. 2163). To investigate this issue, Brown et al. recruited 18 participants who traveled to four sites to complete multiple runs of a working memory task. The results of their study indicate that, generally, between-site reliability of BOLD data can be good to excellent. However, they also indicate that "acquiring highly reliable data requires robust activation paradigms, ongoing quality assurance, and careful experimental control" (p. 2163).

Altogether, the sample studies discussed here indicate that BOLD data is generally fairly stable in test-retest paradigms. Yet, depending on multiple factors (e.g., experimental task and design, the investigated theoretical construct and its neural correlates, or the MRI scanner) reliability of BOLD data might not be ideal[7]. In addition to fMRI investigations, EEG studies have also turned out to play a significant role in NeuroIS research; see the papers by Léger et al. (2014) and Vance et al. (2014) in this special issue, or several papers in a *Journal of Management Information Systems* special issue on NeuroIS (Vol. 30. No. 4). Consequently, reliability of EEG data is also an important issue in NeuroIS research.

---

[5] Note that the instrumenter instructs the subject to correctly position in the MRI scanner and how to behave properly during the measuement session (e.g., avoid head movements). Moreover, the researcher makes judgmental choices about parameter settings and processing algorithms (see, e.g., Huettel, Song, & McCarthy, 2009). It follows that the instrumenter or researcher may indirectly affect measurement reliability. We refer to this issue in Section 5.5.

[6] For details on intraclass correlation coefficients, see, for example, Koch (1982) and Müller and Büttner (1994).

[7] In this context, we refer the reader to another interesting study on test-retest reliability of functional transcranial Doppler ultasonography by Stroobant and Vingerhoets (2001); this study basically confirms the results found in the fMRI domain.

What is known about the reliability of EEG data has also mostly been derived from test-retest paradigms, and such studies have been carried out for decades. For example, Salinsky Oken, and Morehedb (1991) studied test-retest reliability in EEG frequency-analysis and found correlation coefficients between 0.84 to 0.92, with lower values when the time between the measurements was 12-16 weeks rather than in minutes. Later, based on different EEG paradigms, several other scholars reported test-retest correlation coefficients in the range 0.7 to 0.9, and sometimes even higher (e.g., McEvoy, Smith, & Gevins, 2000; Tervaniemi et al., 1999; Näpflin, Wildi, & Sarnthein, 2007, 2008). Altogether, these results suggest that EEG recordings are generally reliable (both if captured during rest and during task execution). Yet, as already discussed for fMRI, several factors (e.g., experimental paradigm or measurement equipment) might affect reliability. Moreover, it has to be considered that reliability might vary as a function of the specific aspects or components studied in EEG research. For example, one study (Thesen & Murphy, 2002) found that reliabilities were higher for latency than for amplitude. Another study (Fabiani, Gratton, Karis, & Donchin, 1987) reports a test-retest reliability of approximately 0.8 for the P300 amplitude.

Regarding physiological measures related to ANS activity, we found that research indicates that test-retest reliability of heart rate variability (HRV) is considered good (e.g., Guijt, Sluiter, & Frings-Dresen, 2007; Mukherjee, Yadav, Yung, Zajdel, & Oken, 2011). Moreover, test-retest reliability of several electrodermal activity measures (e.g., NS.SCR or SCL) has also been studied; correlations are as follows (Dawson, Schell, & Filion, 2007; Schell, Dawson, Nuechterlein, & Subotnik, 2002): in the range of (i) .40 to .75 for NS.SCR frequencies, (ii) .40 to .85 for SCL, and (iii) .30 to .80 for the number of SCRs elicited by repeated stimuli. Importantly, stability of temporal aspects of these measures (e.g., latency or rise time) can be lower (Dawson et al., 2007; Schell et al., 2002).

In general, we recommend, if appropriate in the context of a specific study, that NeuroIS researchers capture multiple scores of the same construct to aggregate the findings by calculating an average. Random sources of error, potentially masking the true score, may cancel out by this simple procedure. Strube and Newman (2007) confirm this view: they write that, “if we average more and more observations, each with its own random error source but measuring the same true score, then the odds of the error canceling out keep improving … provid[ing] a more reliable measure” (p. 792).

5.2. Validity

Validity indicates the extent to which a measurement instrument measures the construct that it purports to measure. Importantly, there is no one true validity. Rather, several different types of validity exist (see, for example, an introduction in Kerlinger and Lee (2000), Chapter 28). In this section, we focus on validity related to data collection, referred to as “instrumentation validity” (e.g., Cook & Campbell, 1979; Straub et al., 2004)[8]. Specifically, we discuss content validity and construct validity in the context of NeuroIS research and physiological measurement.

Content validity concerns the issue of drawing a representative measure, or a representative set of measures, out of all possible measures for a given construct. For example, an IS researcher investigating “user emotion” could use one of several different measures to capture this construct, or a combination of measures. Relevant measures include, but are not limited to, survey- or behavior-based measurement, multiple physiological measures such as activity in specific brain areas (e.g., limbic structures), ANS activation (e.g., heart rate, blood pressure, skin conductance, pupil dilation), and muscle tension changes in specific face areas (e.g., Corrugator supercilii: related to smiling, Zygomaticus major: related to frowning). A study on IS use patterns (Ortiz de Guinea & Webster, 2013), for example, operationalized user emotions based on two dimensions; namely, affective valence (survey measurement) and physiological arousal (ANS measurement).

But how do researchers know whether they have appropriately chosen the measure, or set of measures, so that the essence of the construct is captured well? It is common practice in survey research that content validity is established through literature reviews and expert or panel judgments

[8] Examples of other important types of validity are: internal validity (i.e., ruling out rival hypotheses), statistical conclusion validity (i.e., statistical inference), and external validity (i.e., generalizability) (e.g., Straub et al., 2004).

(Straub et al., 2004)[9]. However, to the best of our knowledge, NeuroIS studies often do not discuss content validity. Thus, the degree of certainty is low with which the researchers know whether they have representatively drawn from the available set of measurement possibilities. Therefore, we recommend that future NeuroIS research projects review the literature more extensively to identify a larger number of potential measures for the construct(s) at hand. With respect to physiological measurement of IS constructs, it is not unlikely that measures pertaining to different analytical levels (e.g., hormonal, CNS, or ANS) may represent the construct as has been shown in context of stress and technostress (for details, see Riedl, 2013). However, we emphasize that measures on different analytical levels (not only on different physiological levels, but also on different conceptual levels, such as cognitive, behavioral, or physiological) may tap into different aspects (dimensions) of a construct. Thus, IS researchers should not generally expect that different measures in a construct domain are substitutes; rather, in many cases, they may be complements; see, for example, the study by Tams et al. (2014) in this special issue for further details.

Construct validity concerns the issue of whether the measures chosen by the researcher (i) go well together (convergent validity) so that they capture the essence of the construct and (ii) diverge from measures capturing distinct constructs (discriminant validity)[10]. Methods such as the multi-trait multi-method (MTMM) matrix and structural equation modeling can be used to assess convergent and discriminant validity of multiple methods. However, a researcher using a single physiological indicator to measure a construct would not be able to assess convergent and discriminant validity (e.g., a scholar who exclusively uses SCL to measure user arousal in a human-computer interaction task). However, as we indicate in the introductory part of Section 5, many IS constructs may be better represented by a combination of measures because IS research is primarily concerned with complex constructs, several of which comprise both cognitive and emotional dimensions (e.g., trust, see, e.g., a review by Riedl & Javor, 2012). Thus, like survey studies, research using physiological measures should also be concerned with construct validity issues.

In this context, the following fundamental question has to be raised: are the data the outcome of true scores or artifacts of the chosen instrumentation? If constructs are valid in the sense of construct validity, relatively high correlations between measures of the same construct based on different measurement tools can be expected (Cambell & Fiske, 1959; Cronbach, 1971). However, research indicates that the demonstration of construct validity is often hampered by two phenomena (e.g., Strube & Newman, 2007, p. 805): (i) a measure, or set of measures, may only capture part of the construct (referred to as “incomplete representation of a construct”), or (ii) a measure, or set of measures, may represent two or more constructs (referred to as “representation of multiple constructs”).

In case (i), the consequence is that important aspects of the constructs are missed and hence not included in an experimental manipulation. Thus, failure to find a relation between measures is—often wrongly—interpreted as a missing relationship between constructs. As an example, workload (i.e., that portion of a person’s limited capacity actually required to perform a particular task) is not fully captured based on self-reports alone (O’Donnell & Eggemeier, 1986). Thus, measuring workload based on self-reports alone, without considering performance-based and/or physiological measures, is problematic. In case (ii), the consequence is that a researcher cannot be certain that their theoretical explanation is correct. Rather, alternative explanations might also be plausible. As an example, SCL does not necessarily reflect attention; rather, it might also reflect motivation, among other constructs (Strube & Newman, 2007). Thus, a researcher who finds an increase of SCL as a consequence of an experimental manipulation should not blindly conclude that this manipulation has caused altered levels of attention. Rather, it is also possible that the manipulation affected other constructs, such as motivation.

Based on Strube and Newman (2007, p. 805), we conceptually summarize the phenomena of both “incomplete representation of a construct” (left) and “representation of multiple constructs” (right) in Figure 5. Note that for readability reasons, we kept Figure 5 as simple as possible. Specifically, we depict the threats to construct validity based on (i) only two measures rather than more than two measures (left) and (ii) only two constructs rather than more than two constructs (right). In reality,

[9] Statistical procedures have also been suggested to formally test content validity (e.g., Lawshe, 1975).
[10] Straub (1989) has tellingly labeled construct validity as the “meaningfulness of constructs as measured“ (p. 151).

most IS constructs, at least those implying a high level of complexity, need to be operationalized based on more than two measures, and many physiological measures may represent dimensions of different IS constructs.

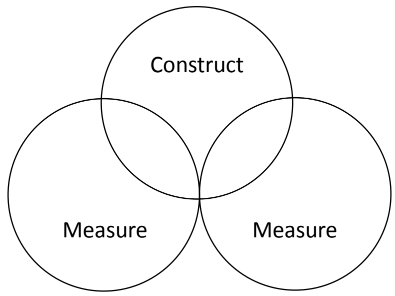

Figure 5. Threats to Construct Validity: Incomplete Representation of a Construct (left) and Representation of Multiple Constructs (right)

Ortiz de Guinea, Titah, and Léger (2013) evaluate IS construct validity based on the MTMM matrix in the context of NeuroIS research. Specifically, they used self-report and physiological instrumentation to measure three IS constructs (engagement, arousal, and cognitive load) in two different experimental settings (instrumental and hedonic tasks). They measured the self-reports with Likert-type instruments and manikins[11]. Physiological measurement was based on EEG and EKG. Without going into the details of the study results, the findings suggest that more primitive perceptual IS constructs (here arousal) are less affected by mono-method bias (i.e., a bias occurring when the measurement technique introduces systematic variance into the measures) if compared to more complex perceptual constructs (here engagement and cognitive load). One major implication of this result is that, whenever an IS researcher studies a complex construct (an attribute that holds true for many IS constructs), the use of one measure only likely puts validity at risk.

In their paper, Ortiz de Guinea et al. (2013) also argue that "neurophysiological measurement appears to have less measurement error due to their direct nature"; however, they also indicate that "it may also have less construct and content validity" (p. 841). The major argument they provide for why neurophysiological measurement in IS research might have lower validity is that physiological measures are often more narrowly focused, and hence it is difficult to tap into the entire construct space (note that an implicit assumption here is that IS constructs are complex). As a rule of thumb: content validation is less difficult when the "distance" from the construct to the measure is "short" (e.g., construct: arousal, measure: SCL). However, when the construct is more complex, identifying an appropriate measure, or set of measures, is more difficult (e.g., construct: workload, measures: subjective, performance-based, and physiological; O'Donnell & Eggemeier, 1986).

In endocrinological research, the concept of accuracy is closely related to validity. Accuracy is "the ability of the assay to measure the true concentrations of the analyte in the samples being tested" (Schultheiss & Stanton, 2009, p. 27). Table 3, based on Schultheiss and Stanton, summarizes standard procedures used in endocrinological research to calculate accuracy of hormone measurements, and provides an example. NeuroIS scholars using endocrinological measurement should consider accuracy in their research.

[11] Manikins are images that represent how participants felt. Scale: five manikins displaying five states from completely calm to completely excited (Ortiz de Guinea et al., 2013, p. 842).

Table 3. Accuracy in Hormone Research	
Definition	The ability of the assay to measure the true concentrations of the analyte in the samples being tested.
Procedure	Control samples (CS) with known amounts of the analyte are included into the assay, which are then compared to the amount of analyte estimated (AE) by the assay.
Metric	Accuracy = AE/CS×100. Coefficients between 90% and 110% indicate good accuracy.
Example	Testosterone: CS = 100 pg/ml, AE = 95 pg/ml, Accuracy = 95/100×100 = 95%, Source: Schultheiss and Stanton (2009, p. 27).

Generally, in order to measure hormone concentrations in a valid way, several potential confounders must be considered in the development of both task instructions and experimental design. For example, Riedl et al. (2012, p. 65) indicate in the methods section of their paper on cortisol measurement in the context of technostress:

> *All participants were given instructions not to drink anything containing alcohol or caffeine, nor to do physical exercises from 7.00 p.m. on the day before their participation. Moreover, they were instructed not to eat and drink anything except water within two hours prior to their participation. Because cortisol levels in humans naturally decrease in the morning but are relatively stable in the afternoon … experimental sessions were conducted between 2.00 p.m. and 6.00 p.m.*

We conclude that extensive knowledge on the determinants of the concentration of a specific hormone is necessary, which includes those related to individual differences (e.g., sex or age), in order to be able to plan and execute a research study with reliable and valid measurement.

5.3. Sensitivity

Sensitivity is a property of a measure that describes how well it differentiates values along the continuum inherent in a construct. Because a measure, by definition, must be capable of assuming at least two values, minimal sensitivity is given when two states of a construct can be discriminated. For example, it is a well-established fact that skin conductance level (measure) is an indicator for arousal (construct). Because skin conductance is capable of both distinguishing high from low levels of arousal (two classes), and also different levels of arousal (on a continuous scale rather than in discrete classes), skin conductance level can be said to be fairly sensitive for the measurement of arousal.

Note that sensitivity of a measure should be harmonized with a research study's requirements. Thus, always choosing the instrument with the highest possible sensitivity to discriminate possible values of a measure is nonessential. Riedl, Kindermann, Auinger, and Javor (2013), for example, investigated whether male users exhibit higher levels of stress than female users in cases of system breakdown during the execution of a human-computer interaction task under time pressure. In this experiment, measurement of stress was based on skin conductance level and considered as sufficiently sensitive to investigate the research question. Citing related work, they write in their paper that skin conductance:

> *was used as a stress indicator because it reliably reflects activity of the sympathetic division of the autonomic nervous system … the part of the human nervous system that is active during perception of arousal and stress …* [t]*his measure is a well-established stress indicator in the human-computer interaction domain* (p. 3).

For the purpose of their experiment, it would have also been possible to use several alternative measures to quantify arousal and stress, including other ANS measures (e.g., EKG: heart rate variability) or endocrinological measures (e.g., hormone assessment based on saliva samples: cortisol) (for details, see Boucsein & Backs, 2000). However, because selection of a particular instrument and measure is affected by multiple criteria (and not only by sensitivity), Riedl et al. (2013) chose skin conductance level primarily because it is less intrusive than the other measures (intrusiveness is explained below in more detail), and because their study was focused on sex differences, implying that interpreting results based on hormone measurement would have been highly complex due to the female menstrual cycle.

Generally, a major methodological question for NeuroIS researchers is whether specific physiological measures (e.g., skin conductance level), including its specific features (e.g., skin conductance response amplitude, or nonspecific skin conductance response frequency), can distinguish at least two states (high, low) of an IS construct, and, if so, whether it can even make a distinction on a higher level of granularity. Thus, an important element of a NeuroIS research methodology is a taxonomy mapping IS constructs and corresponding physiological measures and features in consideration of sensitivity. The taxonomies that Boucsein and Backs (2000) provide may serve as examples (see pages 9-21).

5.4. Diagnosticity

Diagnosticity is a property of a measure, and describes how precisely it captures a target construct as opposed to other constructs. Thus, if a physiological measure represents only one construct, but not other constructs, its diagnosticity is maximal. However, a perfect one-to-one relationship between physiological measures and psychological constructs does not exist (Cacioppo & Tassinary, 1990). Rather, one physiological measure is often related to multiple constructs (Figure 6, left)[12]. Thus, 100% diagnosticity is theoretically possible, but it is very unlikely that it can ever be established in reality in psychophysiological research. With that said, researchers should strive for a high degree of diagnosticity (Figure 6, right). Note that whether or not a specific degree of diagnosticity is sufficient depends on the research context.

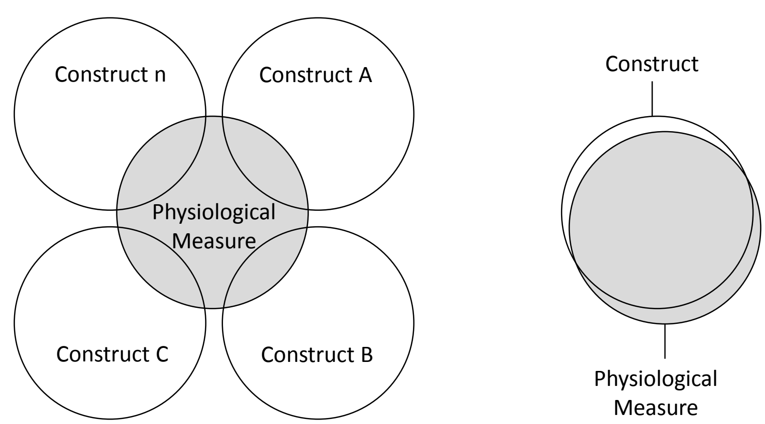

Figure 6. The Concept of Diagnosticity: Low Degree of Diagnosticity (Left) and High Degree of Diagnosticity (Right)

Diagnosticity may also refer to the capability of an instrument to discriminate different subcomponents of a construct, and is hence related to validity. For example, O'Donnell and Eggemeier (1986) investigated workload, the "portion of the operator's limited capacity actually required to perform a particular task" (p. 42-2). Based on the fact that the human processing system can be described as a series of independent resources that are not interchangeable, O'Donnell and Eggemeier decompose workload into different subcomponents. For example, the execution of a human-computer interaction task requires perceptual and central processing and motor commands. These different subcomponents of workload draw from separate resources. The question is now whether a measure indexes workload across the entire processing system including all types of resources, or whether a measure can diagnose the involvement of a specific subcomponent or resource. O'Donnell and Eggemeier (186) write:

[12] Note that the left part of Figure 6 is a more complex representation of the right part of Figure 5. However, the illustrated issue is similar.

> [P]*upil diameter…and some subjective rating scales…appear to index workload across the entire processing system. With such measures, it would not be possible to diagnose which type of resource or capacity (e.g., perceptual versus motor output) had been affected, although an overall assessment of workload would still be possible. On the other hand, the event-related brain potential* [a specific EEG measure]*…and some secondary tasks* [tasks which have to be performed by a subject in addition to the primary experimental task] *… show a greater degree of diagnosticity in that they appear to be maximally sensitive to particular types of resource/capacity expenditure. Use of such measures would permit more precise localization of the source of an overload, although they could be insensitive to loading in unmeasured resources* (p. 42-4).

Thus, one important task in future NeuroIS research is the decomposition of a construct into subcomponents in order to distinguish measures that index a construct across its entire space from measures that are diagnostic of a specific subcomponent of the construct.

5.5. Objectivity

Kerlinger and Lee (2000) write in their seminal textbook on research methodology in the behavioral sciences: "[T]he checks used in scientific research are anchored as much as possible in reality lying outside the scientist's personal beliefs, perceptions, biases, values, attitudes, and emotions. Perhaps the best single word to express this is *objectivity*" (p. 7, italics in original). However, they also write in the same book that "[o]bjectivity, a central and essential characteristic of scientific methodology, is easy to define but evidently hard to understand" (p. 708). In this section, we shed light on the meaning of objectivity in the context of NeuroIS research. Here, we define objectivity as the extent to which research results are independent from the investigator and reported in a way so that replication is possible.

In a NeuroIS research agenda, a group of fourteen scholars motivated their paper in the abstract with the following words: "There is heated interest now in the social sciences in capturing *presumably objective* data directly from the human body, and this interest in neurophysiological tools has also been gaining momentum in IS research" (Dimoka et al., 2012, p. 679, italics added)[13]. The adverb "presumably" is important because research related to the material world (here Neurobiology, which is, in turn, largely based on physical and chemical principles) is usually considered as "100% objective". However, while "neurobiological signals" might be considered as more objective than perceptual data (reported in surveys) and human behavior (because they are more independent of individual thought and often perceptible by all observers in an identical way, typically based on sensors such as electrodes), we claim that "neurobiological research" is not necessarily objective.

Our framework in Figure 2 shows that data collection and data analysis are important aspects of a methodology in NeuroIS research. While a comprehensive discussion of the collection and analysis of physiological data is beyond the scope of this paper (please see, for example, volumes such as Kantz, Kurths, & Mayer-Kress, 1998), we discuss selected aspects related to these two basic research activities, and do so from the perspective of research objectivity. Moreover, we emphasize that objectivity is a concept closely related to replication. Straub et al. (1994) write: "Study results are objective and in such a form that other researchers could replicate the work" (p. 34). Thus, to understand objectivity it is important to understand that (i) data collection, (ii) data analysis, and (iii) application of physiological methods and corresponding results should be reported in ways that allow for replication. We reflect on important aspects in these three domains in the following paragraphs, and emphasize that further contributions in these domains are urgently needed in NeuroIS research.

Data collection is the process of measuring the values of variables, and the objective of this process is to gather evidence which is then used as input for data analysis, and hence is the basis of credible answers to the research question at hand. In research using physiological measures, several individual steps have to be planned and executed. The quality of planning and execution, among other factors, significantly affects a study's objectivity. Assuming that the study design and tasks have been prepared and subjects have been recruited, data collection comprises the following major steps:

---

[13] Note that objectivity of neurobiological measures is referring to independence from the subjectiveness of the subject, rather than the investigator.

physically setting up the measurement instruments and computer systems, greeting the participants, attaching sensors to participants, testing and calibrating measurement instruments, announcing the task instructions to participants, executing experimental sessions, removing the sensors, and debriefing the participants (note that complementary collection of survey data can take place before, during, and/or after collection of physiological data)[14].

Physiological reactions of humans in IS contexts (e.g., human interaction with computers) are usually measured by sensors placed on the body surface, even though the bodily reaction actually occurs “in” the body. This external recording of physiological signals causes distortions, often referred to as noise in the scientific literature. Generally, in physiological research, noise is “any other phenomena observed in the data apart from the signal(s) of interest for the investigator” (Gratton, 2007, p. 848). A measure reflecting the ability to distinguish signals from noise is the “signal-to-noise ratio”; the lower this measure, the more difficult it is to identify the signal. In other words, higher signal-to-noise ratios are desirable. This explains why the increase of the signal-to-noise ratio is an important task in the processing of physiological signals, primarily realized by amplifying the signal, reducing the noise, or both. Techniques, procedures, and devices that reduce the amount of noise present in the data are called “filters”. When the signal-to-noise ratio is large (e.g., >3:1, Gratton, 2007, p. 849), a feature of interest (e.g., the amplitude of a signal) can be relatively easily distinguished from noise, often by simple visual inspection of the data. However, when the level of noise is high, filtering and/or pattern recognition algorithms are required to attribute changes in the physiological signals to experimental manipulations.

Generally, noise obscures the true values of a physiological signal. For example, measuring heart rate is often based on sensors attached to the chest area. Thus, at least two “layers of noise” exist in this simple example. First, the signal coming from the heart muscle must reach the skin (layer 1). Second, there exists the contact area between the skin and the sensor (layer 2). Thus, individual differences in body tissue (layer 1) and sensor (electrode) material and electrolytes (layer 2) directly affect the noise level. Moreover, cables, connectors, and electromagnetic emissions may affect signal quality. To minimize such distortions, measurement instruments often have a first amplification device close to, or directly in, the sensor (Luczak & Göbel, 2000). Also, note that atlases exist for the placement of sensors. Depending on the specific instrument at hand (e.g., fEMG, EKG, EEG), researchers should strictly follow the placement instructions provided in method guidelines. It is obvious that placement of electrodes on the body surface affects noise. Thus, differences in placement of electrodes may affect research findings and thus place test-retest reliability and successful replication of results at risk. Moreover, exact descriptions of filtering techniques and the applied pattern recognition algorithms are indispensable, otherwise a study does not deserve the attribute “objective”.

Another important aspect in the processing of physiological signals is “data sampling”; that is, the reduction of a continuous signal to a discrete signal. The unit of signal frequency used is Hertz (Hz)[15]. The Hz is equivalent to cycles per second. Importantly, the temporal resolution of a sensor signal directly determines parameter accuracy, which the following example demonstrates (Figure 7). Figure 7 shows an example of aliasing due to insufficient sampling rate, based on Pizzagalli (2007, p. 64). The vertical axis shows the amplitude of an EEG signal in microsolts (µV), and the horizontal axis shows the time in milliseconds (ms). A 10 Hz sine waveform (illustrated on the top) is digitized at two different sampling rates. As the left side illustrates, the sampling rate (50 Hz) is greater than twice the waveform frequency (A), which results in an appropriate digital representation of the analog signal (B). In contrast, as illustrated on the right side, the sampling rate (16 Hz) is less than twice the waveform frequency (C), leading to a false (aliased) representation of the analog signal (D). Considering this example, it is clear that under sampling leads to an irreparable lower frequency component in the digital signal, and hence data sampling may significantly affect data accuracy, an important precondition for objectivity.

[14] Note that these steps are different in case of fMRI and similar tools (e.g., PET). Here, measurement instrumentation is already set up in specific rooms (i.e., equipment is not mobile) and no sensors have to be attached. Rather, subjects have to be placed in the fMRI machine in a supine position.

[15] Hz is named by Heinrich R. Hertz (1857-1894), a German physicist who first conclusively proved the existence of electromagnetic waves.

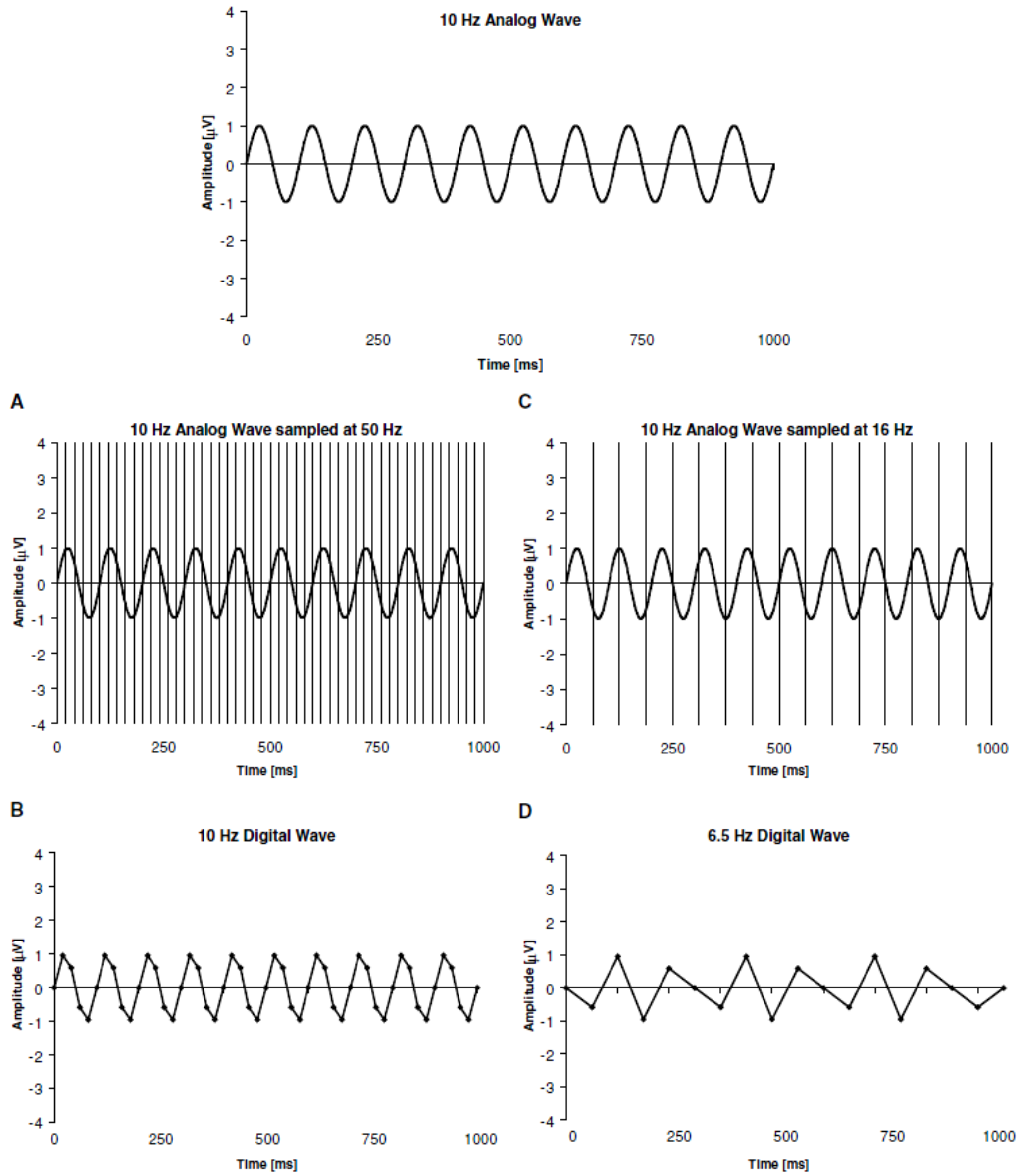

Figure 7. Example of Aliasing due to Insufficient Sampling Rate (Source: Pizzagalli, 2007, p. 64)[16]

We indicate above that filters are procedures that reduce the amount of noise present in the physiological data. Low-pass filtering reduces noise in raw signals (i.e., reduces the amplitude of signals with frequencies higher than a specific cutoff frequency). However, this occurs at the cost of temporal resolution. Generally, temporal resolution, which describes the time that a physiological signal needs to respond to the onset of a specific stimulus, is an important aspect in psychophysiological research. For example, electrodermal response to discrete events can be expected within a time interval of 3 seconds (Dawson et al., 2007). Gratton (2007) indicates that measures with a short time constant (usually measured in milliseconds) have a good temporal resolution (e.g., ERPs, MEG, EMG, or eye movements), and therefore the effects of two different (experimentally manipulated) events occurring in rapid succession can be distinguished, whereas

[16] Aliasing occurs when a signal is sampled at a rate that is too low. Further graphical representations of aliasing can be found, for example, in Gratton (2007, p. 838) or Luczak and Göbel (2000, p. 84).

measures with a long time constant (e.g., usually measured in seconds or minutes) have a poor temporal resolution (e.g., EDA, fMRI, PET, or specific hormones), and hence events in rapid succession cannot be distinguished easily. It is critical that NeuroIS researchers choose their measurement instrument and the sampling rate and filters deliberately. Specifically, to avoid problems, "fast phenomena" (e.g., attention) should be examined with appropriate tools (e.g. EEG), a high sampling rate, and careful use of low-pass filtering. Examples for problematic sampling rates are when (i) EEG is measured at a rate <20 Hz, (ii) EKG is measured at a rate <2 Hz, or (iii) respiration is measured at a rate <0.5 Hz (Gratton, 2007, p. 838). As a rule of thumb, the sampling rate needed in a NeuroIS study depends on the research question at hand (including the phenomenon or construct, respectively, and the experimental design) and the temporal occurrence of the physiological signal. However, this temporal occurrence is constrained by natural limitations. For example, the heart rate of healthy humans is between 60-100 beats per minute (BPM) at rest, and that of athletes at the elite level can be at 30 BPM (at rest) and at 220 BPM (at all-out effort to the point of exhaustion).

To sum up, collection of physiological data (e.g., where to exactly attach sensors on the subject's body), preparation of this data for subsequent analysis (e.g., filtering), and how both collection and preparation are reported in a manuscript are activities that may significantly affect objectivity, and hence whether a study can be replicated. Next, we discuss selected aspects related to data analysis in the context of research objectivity.

Data analysis is the process of inspecting, cleaning, transforming, converting, aggregating, and modeling data in order to develop useful information. This information, in turn, is the basis for drawing conclusions (e.g., rejecting a hypothesis). Unlike survey data, physiological raw data (i.e., the original signals captured by the instrument) are typically too complex to analyze without data reduction. Thus, the extraction of specific features of the data is important to handle the complexity. These features (e.g., amplitude, temporality, or frequency) are the input for statistical analyses, particularly hypotheses testing. Generally, the variety of data analysis techniques is large, and selecting a specific technique depends on multiple factors ranging from level of measurement (nominal, ordinal, interval, ratio) to the number of investigated variables (univariate, bivariate, multivariate), among others.

Jennings and Gianaros (2007) report on usage frequency of different analysis methods in psychophysiological research (1 = highest frequency and 6 = lowest frequency): 1) ANOVA and ANCOVA, 2) t-tests, 3) multiple regression and correlation, 4) multivariate analysis of variance, 5) nonparametric analysis, and 6) factor analysis, principle components analyses, multidimensional scaling, path analysis, and structural equation modeling. A majority of these methods are well-known in the IS field, and the mainstream IS PhD student is trained in the application of most of these methods. Thus, while collecting and preprocessing (e.g., filtering) physiological data often has unique characteristics, if compared to traditional IS methods, most IS researchers are familiar with the analysis methods used in psychophysiological research, and this circumstance might positively affect the development of NeuroIS. Moreover, note that a considerable number of IS scholars are well trained in applying structural equation modeling (SEM). Because this method is not yet well established in psychophysiological and brain research, IS experts in SEM may provide methodological contributions to the Cognitive Neuroscience literature[17]. Thus, IS research should not only be considered as a "passive consumer" of methodological knowledge from physiology, but rather also as an "active producer" of such knowledge. We refer interested scholars to an article by Boucard, Marchand, and Nogues (2007) on the reliability and validity of SEM applied to neuroimaging data to get a first impression about this research domain.

In the following paragraphs, we discuss selected issues related to analyzing physiological data. Most of these issues are relatively independent from the used instrument. Imagine the following research question: does system breakdown in a human-computer interaction task increase users' physiological activation and stress? We indicate in Section 2 that activity of the sympathetic division of the ANS is a consequence of stress perceptions, and is related to a number of indicators, such as an increase in EDA, heart rate, pupil dilation, muscle tension, and elevation of stress hormones. Like in most other psychophysiological research situations, in this example, the chosen measure (let us assume that it is

[17] We thank David Gefen for corresponding reflections during a panel discussion at the Gmunden Retreat on NeuroIS 2012.

EDA; for an empirical contribution in this context, see Riedl et al., 2013) is intended to reflect a change in a biological variable from a level that exists prior to an experimental manipulation (e.g., system breakdown error message in the form of a pop-up on the screen) to a level that exists after the manipulation. The level of the biological variable before the introduction of the manipulation is called baseline, and, after stimulus onset, a researcher expects the physiological response (see Figure 8; note that we refer to the activities before the beginning of a task later in this paper).

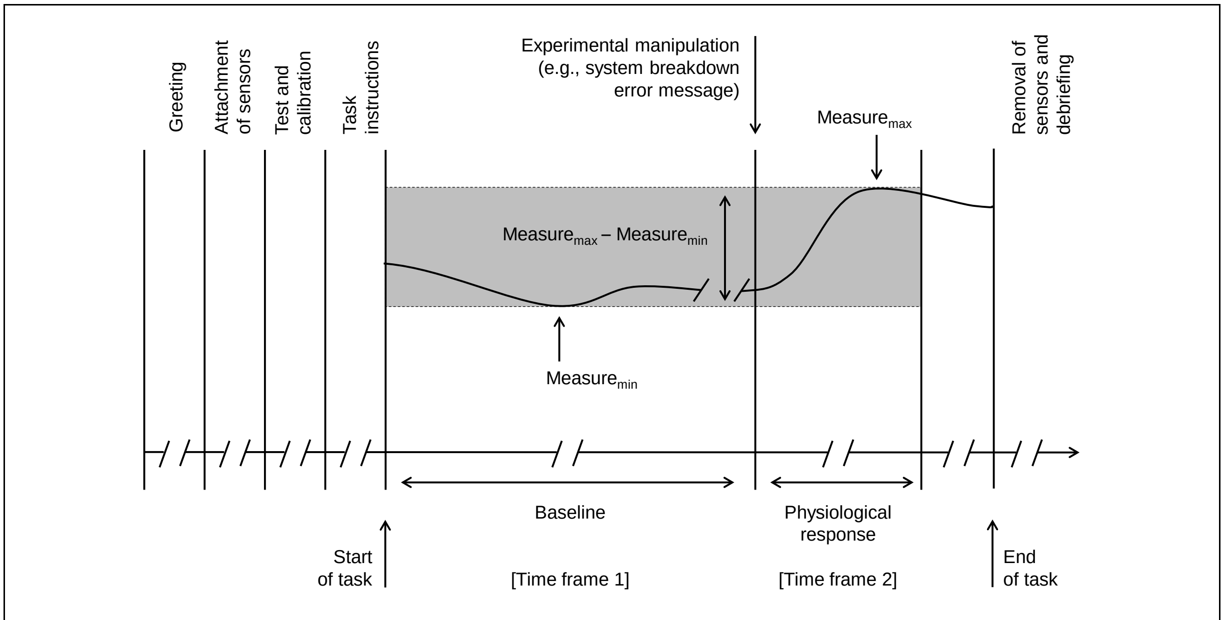

Figure 8. Illustrative Experimental Protocol of a NeuroIS Study

But why is it important to have baseline measurements in psychophysiological research? Almost all physiological systems are permanently active (e.g., heartbeat, respiration, or brain). Thus, physiological activity exists prior to an experimental manipulation. However, individuals physiological activity may significantly differ during rest (i.e., before the manipulation) if compared to activity during stimulus perception and/or task execution, and hence the "change of a measure" that results from the experimental manipulation is the focus of analysis in most psychophysiological studies. Importantly, not only does resting physiology change as a function of subject (e.g., sex, age, health status, or unspecific differences), but several other factors may also affect baseline activity, including time of measurement or posture. For example, several hormones follow a circadian rhythm, and therefore holding measurement time constant across all subjects, or entering this factor as a control variable into data analysis, is critical. As another example, it is well known that posture (e.g., laying, standing, sitting) affects heart rate via complex ANS and cardiovascular processes. For these and many other reasons (see, for example, standard textbooks in Human Neurobiology and Cognitive Neuroscience, such as Gazzaniga et al., 2009), physiological measures are usually expressed as changes with reference to a baseline level, which results in normalized values. A simple procedure to normalize each measurement point (Measurement$i$) is based on the following formula (see Figure 8):

$$(\text{Measurement}i - \text{Measurement}_{min}) / (\text{Measurement}_{max} - \text{Measurement}_{min}).$$

Thus, a major goal of normalization is to balance inter-individual differences in physiology. In this context, it is also important that NeuroIS researchers develop strategies for the handling of outliers; that is, data points that deviate (too) significantly from the rest of the data set and that are likely the result of data collection errors, possibly due to malfunction of sensor equipment or software tools. These strategies should consider general guidelines on the handling of outliers (e.g., Barnett & Lewis, 1994).
Figure 8 depicts the basic temporal order of activities in experiments based on physiological measurement; namely, that once a task has been started, baseline measurement is taken (time frame 1), followed by stimulus onset, and measurement of the physiological response (time frame 2). A

fundamental question in data analysis concerns the length of both time frame 1 and time frame 2. Obviously, variation in length may affect research results. While the definition of time frame 1 usually depends on the goal of a research study (and typically ranges from a few minutes to 10-15 minutes in experiments based on measurement of ANS activity), the definition of time frame 2 must consider the physiological properties of the investigated signal. For example, humans‘ electrodermal response to discrete events (e.g., system breakdown) can be expected within a period of 3 seconds (Dawson et al., 2007). Thus, a reasonable time frame 2 would be 3 seconds (to investigate whether system breakdown significantly affects activation and stress, if compared to a situation in which the execution of a human-computer interaction task is not disturbed by system malfunction).

Generally, it is important for NeuroIS researchers to know as exactly as possible the response function of a specific physiological signal that results from the perception of: (i) a short-term event or (ii) a block of multiple consecutive events. Important information with respect to the physiological signal includes, but is not limited to: onset latency after stimulus perception, time to peak, amplitude, and time to return to baseline. As an example, the BOLD hemodynamic response (HDR)[18] which forms the basis of fMRI data analysis takes, in case of perception of a short-term discrete event, the following stereotyped form: initial dip (likely the result of initial oxygen extraction before the later overcompensatory response, see Huettel et al., 2009), a subsequent increase to a peak around 5 seconds after stimulus onset, followed by a return to baseline and subsequent undershoot at around 12-15 seconds; then, after 20-25 seconds after stimulus onset, the signal returns to the original baseline value. Importantly, analysis software in fMRI research uses a default hemodynamic response. However, note that evidence shows that the HDR is not independent from the properties of the stimulus (Huettel et al., 2009). Moreover, research indicates that elderly subjects have a significantly reduced signal-to-noise ratio in the fMRI signal if compared to younger subjects (D'Esposito, Zarahn, Aguirre, & Rypma, 1999). This finding suggests that the coupling between neural activity and the BOLD signal changes with age. While the exact source of this difference is not well understood today, D'Esposito et al. argue that vascular changes rather than neural changes in normal aging might be the source. The implication of this finding is that findings of fMRI studies that compare individuals of different age groups must be interpreted with caution. Generally, normal changes in physiology that result from aging must not be ignored in NeuroIS research, particularly in studies with subjects from different age groups.

With respect to data analysis, another important issue concerns the absolute level of physiological baseline activity. Specifically, the physiological response that occurs due to stimulus perception is not independent from the absolute level of baseline activity. Higher absolute baseline levels of a physiological variable might lead to a limited increase in that variable, and lower absolute baseline levels might lead to a limited decrease. For example, if the resting heart rate of a subject is 140 BPM, its potential to increase as a consequence of stimulus perception (e.g., stress factor such as system breakdown) is not as high as it would be for a subject with a resting heart rate of 80 BPM; conversely, if the resting heart rate of a subject is 40 BPM, its potential to decrease as a consequence of stimulus perception (e.g., relaxing music) is not as high as it would be for a subject with a resting heart rate of 80 BPM (note that healthy adults have heart rate between 60-100 BPM at rest). This principle, known as the "law of initial values", was described comprehensively decades ago (e.g., Wilder, 1967). Researchers have since reconceptualized this law (e.g., Jin, 1992); yet, the basic idea described above is valid, and hence fundamental for Psychophysiology in general and NeuroIS research in particular.

Another aspect related to data analysis and objectivity that we consider as essential concerns the level of measurement; Jennings and Gianaros (2007) write that:

> [D]*ata obtained in different conditions can be interpreted either along ordinal scales, which require using less powerful nonparametric statistics, or along interval scales, which may be analyzed using more powerful parametric statistics* [and] *the basic assumption of interval scales is that differences between intervals at any level of the scale are directly comparable* (p. 855).

[18] The HDR results from a decrease in the amount of deoxygenated hemoglobin present within a voxel (Huettel et al., 2009).

However, the NeuroIS researcher should be aware of the fact that this kind of linearity often does not exist in Psychophysiology. Thus, the difference between the values 5 and 10, for example, does not necessarily have the same significance as the difference between the values 55 and 60. One of the reasons why most psychophysiological measures depart from linearity is that measures usually have a limited range due to feedback mechanisms that counteract extremely high or low values in order to maintain homeostasis, a state of the biological system in which the body is in a stable and constant condition. Cortisol, for example, has a number of adaptive functions in stress situations (e.g., related to cognition, memory, and emotion). However, cortisol also serves the function of counteracting the primary stress response of the body by reducing activation in brain structures that release precursor substances of cortisol (CRH and ACTH; more details on the underlying physiology relevant from an IS perspective is reported in Riedl, 2013).[19] In addition to feedback mechanisms, mechanical constraints may also preclude extreme values (e.g., BPM of the heart).[20]

Despite the complexities related to data analysis and non-linearity, Jennings and Gianaros (2007) indicate that, for most psychophysiological measures, an interval of values exists for which linearity is valid. Consequently, we recommend, if possible, that measurements be collected in experimental conditions in which the physiological measures are in the “linearity range”. As another solution to the problem, Jennings and Gianaros stress that special transformations (e.g., “logit” or “probit” transformations) could also be used to correct special forms of departures from linearity.

Figure 8 shows an illustrative experimental protocol of a NeuroIS study. So far, our discussion has been focused on the middle part of this figure; namely, the aspects related to baseline and response measurement and corresponding data analysis aspects. However, another important aspect closely related to objectivity of a study concerns the activities before the actual beginning of the experimental task, such as greeting, attaching sensors, or announcing the task instructions. Specifically, the “social context of a study” based on physiological measurement (independent from being conducted in the laboratory or in a field setting) might directly affect objectivity. Fridlund and Cacioppo (1986) write:

> *Inattention to subject-experimenter effects may be a remnant of the belief that, unlike verbal and overt behavior, physiological measures are objective and bias-free… The vulnerability of physiological responses to instructional sets…intentional distortion…and social biases…vitiates this notion* (p. 580).

Similarly, Gale and Baker (1981) argue that “[i]n psychophysiological studies, experimenter-subject interactions are particularly important since the procedures may involve bodily contact, partial removal of clothing, skin abrasion, touching, and application and removal of electrodes” (p. 373). We are not aware of systematic examination of the effects of subject-experimenter interaction in NeuroIS research. To close this significant research gap, future research is needed. What we know from informal communication with colleagues working in the NeuroIS field is that studies are often not designed to maximize objectivity. For example, task instructions are often read by the experimenter or by a research assistant (and sometimes even only explained “from memory”). Also, at least in some cases, task instructions are not always read by the same person during all experimental sessions. It is obvious that those practices negatively affect objectivity and may confound research results.

Moreover, evidence shows that observation by, or the presence of others, tends to facilitate performance on simple tasks and impair performance on complex tasks (Fridlund & Cacioppo, 1986). Thus, depending on the facilities where the data are collected (e.g., one-way mirror vs. conspicuously placed camera vs. experimenter is in the room), a study may be more or less objective. Importantly, because those details are often not reported in publications, exact replications are not possible. With respect to the room in which data are collected, it is also important to hold the temperature, humidity, and lighting conditions, among other factors, constant across all experimental sessions. It is obvious that these factors may significantly affect electrodermal activity or pupil dilation, among other measures, rendering data unusable if variance exists across experimental conditions.

[19] CRH = corticotropin-releasing hormone, ACTH: adrenocorticotropic hormone.
[20] We thank Walter Struhal, Linz General Hospital, for this note in the context of the Gmunden Retreat on NeuroIS 2014.

Reflecting on the possible negative effects of social interaction among the experimenter and the subject in the context of psychophysiological research, Cacioppo, Petty, and Marshall-Goodell (1985, p. 288) indicate that both evaluation apprehension (i.e., subjects are apprehensive about being evaluated by the experimenter) and demand characteristics (i.e., subjects try to discern the true purpose of the study and shape their behavior accordingly) can be minimized by employing specific techniques: (i) giving the study participant a false hypothesis, (ii) stressing, based on a cover story, that the neurophysiological mechanisms being examined are not subject to voluntary control, (iii) attaching dummy sensors on areas that lend credence to the cover story, (iv) designing the setting and procedure to minimize the participant's feeling of being scrutinized, (v) increasing experimental realism by using treatments that are sufficiently absorbing, and (vi) reducing the difference in status between the experimenter and participant (e.g., through the establishment of rapport). Cacioppo et al. (1985) tellingly argue that "the nature of the interaction between the experimenter and subject is both a fount of potential biases in psychophysiological research and the source of their solutions" (p. 288). This fact should be kept in mind by NeuroIS researchers in order to maximize objectivity.

Altogether, Section 5.5 shows that a NeuroIS researcher has to make a multitude of decisions with respect to research design, data collection, pre-processing, and analysis. These decisions may affect the corroboration and/or rejection of the research hypothesis. Consequently, in order to consider a NeuroIS study as objective, it is important that authors report details related to study design, data collection, pre-processing, and analysis in their papers. However, there are often so many details that it is virtually impossible to report all methodological facets of a study, neither in the paper nor in an appendix (because journal space is scarce). Solutions to this challenge are that journals provide methodological details of empirical NeuroIS papers on their websites (e.g., in the form of online appendices), or authors provide the details at least during the review process, so that editors and reviewers can check possible effects of methodological aspects on research results. However, the second solution impedes replication, and hence objectivity suffers. Also, note that practices vary significantly in different scientific disciplines. Thus, an active discourse on this issue should be instigated in the IS field.

5.6. Intrusiveness

The word "intrusive" is defined as "annoying someone by interfering with their privacy" ("Merriam-Webster Encyclopedia"). This general meaning is similar to the word's meaning in the context of research methodology. Intrusiveness indicates the extent to which a measurement instrument interferes with an ongoing task, which thereby distorts the investigated construct[21]. In psychophysiological research, intrusiveness is considered as a main criterion for selecting a measurement instrument (e.g., Allanson & Fairclough, 2004; O'Donnell & Eggemeier, 1986). Therefore, intrusiveness must not be ignored in NeuroIS research. Rather, it must have a prominent place in an agenda on research methodology.

In the NeuroIS literature, intrusiveness is primarily discussed as a "current challenge". Riedl et al. (2010a) write in this context that:

> [D]*uring an fMRI experiment, for example, participants are required to lie still on their back within the scanner while their head is restrained with pads to prevent head motion.…Experimental situations in fMRI studies are artificial, because in real life, computer users usually sit in front of their computer in a familiar, comfortable, and quiet environment. Tools to measure psychophysiological responses are less intrusive, as participants usually sit in front of computers in a quiet environment, but still involve the use of sensors attached to the body* (p. 255).

Importantly, this statement indicates two important facets related to intrusiveness. First, intrusiveness has multiple dimensions. Second, the relative significance of these dimensions may vary. Figure 9 shows a graphical conceptualization of intrusiveness of a physiological measurement instrument.

[21] Note that obtrusiveness ("noticeable in an unpleasant or annoying way," Merriam-Webster Encyclopedia) and sometimes also reactivity ("readily responsive to a stimulus," Merriam-Webster Encyclopedia) are used as synonyms in the scientific literature on research methodology. In the latter case, the word "stimulus" does not refer to the stimulus in an experiment; rather, the measurement instrument is meant in the present context.

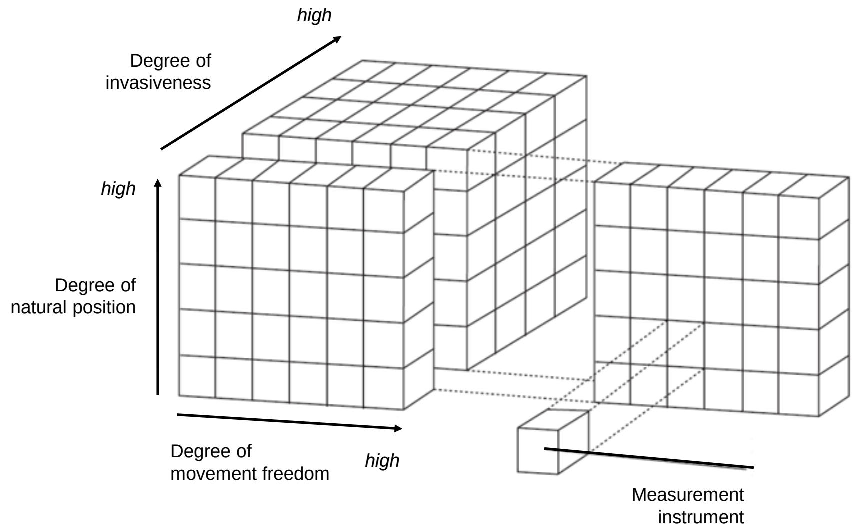

Figure 9. Conceptualization of Intrusiveness of a Physiological Measurement Instrument

In essence, Figure 9 shows that intrusiveness of a measurement instrument can be conceptualized with three dimensions:

1) **Degree of movement freedom:** this factor indicates whether a person is able to move during task execution. As an example, freedom of movement is very low, or even non-existent, in case of fMRI or PET, while it is high in case of mobile and hence wireless devices, such as mobile EDA measurement or mobile eye-tracking.

2) **Degree of natural position:** this factor indicates whether a person is able to carry out a task in a natural position. As an example, if the physiological effects of human-computer interaction are to be studied, then the natural position of interaction is either a sitting position (e.g., PC use) or standing position (e.g., smartphone use). A laying position, however, would be unnatural in most cases in this research context. Accordingly, if the research context was, for example, human-computer interaction, the degree of natural position would be rated as low in case of fMRI or PET because subjects have to lie still on their back in a scanner. However, many other tools, both those related to CNS activity measurement (e.g., NIRS) and ANS activity measurement (e.g., EDA), would score highly on the degree of natural position in the present research context.

3) **Degree of invasiveness:** invasiveness is the extent to which the recording device of a measurement instrument has to be inserted into or attached to the body. For example, if a hormone cannot be assessed reliably in saliva, blood samples must be taken. However, collection of blood samples implies insertion of a venous catheter. Thus, this method is fairly invasive. PET, to state another example, is also a method with a relatively high degree of invasiveness because substances have to be injected intravenously to measure metabolic activity of the brain. Moreover, invasiveness is also determined by the number of sensors to be attached on the body surface. Such sensors, typically electrodes, are used to record physiological activity. Methods such as fMRI, PET, MEG, TMS, or eye-tracking do not use sensors at all, while other methods such as EEG (scalp), NIRS (scalp), EKG (chest or finger), fEMG (face), EDA (hands or feet), or tDCS (scalp) use sensors that are attached at different body locations (as indicated in parentheses). The number of used sensors may vary significantly across methods and within methods. For example, while EDA uses two electrodes, EEG may use, for example, 64 or 128

electrodes (channels). The placement location itself may also affect perceived invasiveness. For example, using fEMG typically requires attaching multiple sensors in the face, while mainstream EDA measurement requires attaching sensors only at the hands or feet (e.g., the non-dominant hand). Consequently it is likely that perceived intrusiveness is higher in case of fEMG than EDA. Empirical validation of this and similar questions is important for the development of NeuroIS research, particularly from a methodology perspective.

Generally, a low degree of intrusiveness exists when 1) the degree of movement freedom is high, 2) the degree of natural position is high, and 3) the degree of invasiveness is low.

Thus, the most desirable position in the cube illustrated in Figure 9 is the right upper corner. However, today, only a very limited number of tools are available that reach this position. For example, bracelets monitoring physiological signals such as heart rate or skin conductance are commercially available, and such devices can be used in human-computer interaction studies, both in laboratory environments and field settings. Generally, technological developments, such as the trend towards increasing miniaturization or the availability of more and more wireless devices, will likely result in less intrusive measurement instruments in the future.

Note that intrusiveness is not the only relevant criterion for tool selection. Rather, selection is based on many other criteria, including spatial and temporal resultion of the physiological signal (see, e.g., Parasuraman & Rizzo, 2008, p. 7) and more pragmatic considerations, such as cost, accessibility, or the knowledge necessary to apply a specific tool. However, NeuroIS researchers evaluating the intrusiveness of their tools in a specific research context may use our conceptualization to classify their measurement instrument. In Figure 9, we hypothetically classify an instrument that has the following characteristics: degree of movement freedom (x-axis: medium), degree of natural position (y-axis: low), and degree of invasiveness (z-axis: low). We leave it to future NeuroIS research to classify all relevant methods and note that positioning of instruments should be based on a clearly defined research context, otherwise evaluating the classification usefulness is difficult, or even impossible.

5.7. Summary

The NeuroIS field contends that physiological measurement is an important complement to the more traditional measurement techniques in IS research. However, application of instruments measuring psychophysiological states and processes comes along with multiple methodological challenges. It is our primary goal to intensify the scientific discourse on six factors closely related to a NeuroIS methodology; namely, reliability, validity, sensitivity, diagnosticity, objectivity, and intrusiveness. We summarize these six factors in Table 4, along with important concepts discussed in the present paper, and emphasize that NeuroIS researchers should carefully give thought to these factors[22]. Obviously, these factors and the ways how the challenges in each domain are addressed significantly determine the methodological quality of a research study, and considering that “conduct of research” (Straub et al., 1994) has been identified as the major criterion for high-quality research, it is clear that disregarding methodological aspects such as those discussed in this paper would be a disservice to the prosperous future development of NeuroIS.

[22] Note that relationships exist among the six factors. Moreover, the six factors are not disjoint categories.

Table 4. Six Major Factors of a NeuroIS Research Methodology and Corresponding Challenges

Factor	Definition	Summary of discussion in this paper
Reliability	The extent to which a measurement instrument is free of measurement error, and therefore yields the same results on repeated measurement of the same construct.	Major sources of measurement error are measurement instrumentation, the experimenter, situational factors, and subject-related factors. Test-retest reliability is the dominant metric to establish reliability in psychophysiological studies. Previous research indicates that test-retest reliability of physiological measures is usually good; yet, it is by far not perfect. Aggregation of findings across multiple measurements may positively affect reliability.
Validity	The extent to which a measurement instrument measures the construct that it purports to measure.	Content validity can be established through literature reviews and expert or panel judgments in NeuroIS research. Construct validity is threatened by two phenomena: (i) a measure may only capture part of the construct, or (ii) a measure may represent two or more constructs. Many IS constructs are complex and should hence be measured on different analytical levels.
Sensitivity	A property of a measure that describes how well it differentiates values along the continuum inherent in a construct.	Physiological measures, including its specific features (e.g., response amplitude), should distinguish at least two states (high, low) of an IS construct. However, many NeuroIS research questions require a distinction on a higher level of granularity.
Diagnosticity	A property of a measure that describes how precisely it captures a target construct as opposed to other constructs.	One physiological measure is often related to multiple IS constructs, and hence maximal diagnosticity can hardly be established in NeuroIS research. Diagnosticity may also refer to the capability of an instrument to discriminate different subcomponents of a construct. Thus, decomposition of a construct into its subcomponents, with the goal to distinguish measures which are diagnostic of a specific subcomponent, is critical.
Objectivity	The extent to which research results are independent from the investigator and reported in a way so that replication is possible.	Unlike survey data, physiological raw data are typically too complex to analyze without data reduction. Thus, the extraction of specific features of the data (e.g., amplitude) is important to handle the complexity. However, procedures related to data collection, pre-processing, and analysis, and how they are reported in a publication, significantly affect objectivity. Thus, both physiological data and NeuroIS research are not objective per se. Moreover, baseline measurement and consideration of the "law of initial values" is important in NeuroIS research. Also, the NeuroIS researcher must be aware of the "non-linearity" of psychophysiological data. The social context of a NeuroIS study (e.g., subject-experimenter interaction) also affects objectivity.
Intrusiveness	The extent to which a measurement instrument interferes with an ongoing task, thereby distorting the investigated construct.	Three major dimensions of intrusiveness are degree of movement freedom, degree of natural position, and the invasiveness of an instrument (e.g., body locations and number of electrodes that have to be attached). In most research situations, a "dominant tool" does not exist, and hence the NeuroIS researcher has to make trade-offs. Thus, in addition to intrusiveness, other selection criteria (e.g., spatial or temporal resolution) have to be considered. Technological developments (e.g., miniaturization, more and more wireless devices) will likely result in less intrusive measurement instruments.

6. Concluding Comments

The genesis of NeuroIS took place in 2007. Since then, a considerable number of IS scholars, and also academics from related fields (e.g., information science), have started to use methods and tools from Neuroscience and Psychophysiology to better understand human cognition, emotion, and behavior in IS contexts. Moreover, using Neuroscience methods and tools may also contribute to the design and development of innovative applications, which neuro-adaptive information systems demonstrate. However, because the NeuroIS field is still in a nascent stage, it is critical that IS scholars become familiar with the methods, tools, and measurements that are used in Cognitive Neuroscience, Neurobiology, and Psychophysiology. Based on a higher degree of familiarity and sound methodological knowledge, IS scholars can better evaluate whether or not a specific method, tool, or measurement is suitable to study a specific IS research question, or may form the basis for the development of a neuro-adaptive information system.

Against the background of the increased importance of methodological discussions in the NeuroIS field, we published a special issue call for papers entitled “Methods, tools, and measurement in NeuroIS research” in 2012. Three papers (Léger et al., 2014; Tams et al., 2014; Vance et al., 2014) were accepted after a stringent review process, in which interdisciplinary review teams evaluated the quality of twenty submissions. These three papers are important steps toward a NeuroIS research methodology. In addition to these three papers, it was our goal to further foster discourse on a NeuroIS research methodology, and to this end, we present this paper. Importantly, our observations during the review process (particularly with respect to the methodological issues raised by the reviewers and associate editors), our own reading of the literature relevant to methodology, and the statements and comments provided by participants of the Gmunden Retreat on NeuroIS during the past years served as input for this contribution. We hope that our discussion instigates future research on a NeuroIS research methodology.

Acknowledgments

We would like to thank all members of the special issue advisory and editorial board and the reviewers for their excellent work.

References

Allanson, J., & Fairclough, S. H. (2004). A research agenda for physiological computing. *Interacting With Computers*, *16*(5), 857-878.

Astor, P. J., Adam, M. T. P., Jerčić, P., Schaaff, K., & Weinhardt, C. (2014). Integrating biosignals into information systems: A NeuroIS tool for improving emotion regulation. *Journal of Management Information Systems*, *30*(3), 247-278.

Bacharach, S. S. (1989). Organizational theories: Some criteria for evaluation. *The Academy of Management Review*, *14*(4), 496-515.

Bagozzi, R. P. (2011). Measurement and meaning in information systems and organizational research: Methodological and philosophical foundations, *MIS Quarterly*, *35*(2), 261-292.

Barnett, V., & Lewis, T. (1994). *Outliers in statistical data*. Chichester: John Wiley & Sons.

Boucard, A., Marchand, A., & Nogues X. (2007). Reliability and validity of structural equation modeling applied to neuroimaging data: A simulation study. *Journal of Neuroscience Methods*, *166*(2), 278-292.

Boucsein, W., & Backs, R. W. (2000). Engineering psychophysiology as a discipline: Historical and theoretical aspects. In R. W. Backs & W. Boucsein (Eds.), *Engineering psychophysiology: Issues and Applications* (pp. 3-30). New Jersey: Lawrence Erlbaum.

Brod, C. (1984). *Technostress: The human cost of the computer revolution*. Reading: Addison-Wesley.

Brown, G. G., Mathalon, D .H., Stern, H., Ford, J., Mueller, B., Greve, D. N., McCarthy, G., Voyvodic, J., Glover,G., Diaz, M., Yetter, E., Ozyurt, I. B., Jorgensen, K. W., Wible, C. G., Turner, J. A., Thompson, W. K., & Potkin, S. G. (2011). Function biomedical informatics research network. Multisite reliability of cognitive BOLD data. *Neuroimage*, *54*(3), 2163-75.

Cacioppo, J. T., & Tassinary, L. G. (1990). Inferring psychological significance from physiological signals. *American Psychologist*, *45*(1), 16-28.

Cacioppo, J. T., Petty, R. E., & Marshall-Goodell, B. (1985). Physical, social, and inferential elements of psychophysiological measurement. In P. Karoly (Ed.), *Measurement strategies in health psychology* (pp. 263-300). New York: John Wiley & Sons.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81-105.

Carter, C. S., Pournajafi-Nazarloo, H., Kramer, K. M., Ziegler, T. E., White-Traut, R., & Bello, D. (2007). Oxytocin: Behavioral associations and potential as a salivary biomarker. *Annals of the New York Academy of Sciences*, *1098*, 312-322.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*(3), 309-319.

Cook, T. D., & Campbell, D. T. (1979). *Quasi experimentation: Design and analytical issues for field settings*. Chicago: Rand McNally.

Cronbach, L. J. (1971). Test validation. In Thorndike, R. L. (Ed.), *Educational measurement* (2nd edn.) (pp. 443-507). Washington DC: American Council on Education.

D'Esposito, M., Zarahn, E., Aguirre, G. K., & Rypma, B. (1999). The effect of normal aging on the coupling of neural activity to the BOLD hemodynamic response. *NeuroImage*, *10*(1), 6-14.

Dawson, M. E., Schell, A. M., & Filion, D. L. (2007). The electrodermal system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (3rd edn.) (pp. 159-181). Cambridge: Cambridge University Press.

Dimoka, A. (2010). What does the brain tell us about trust and distrust? Evidence from a functional neuroimaging study. *MIS Quarterly*, *34*(2), 373-396.

Dimoka, A. (2012). How to conduct a functional magnetic resonance (fMRI) study in social science research. *MIS Quarterly*, *36*(3), 811-840.

Dimoka, A., Banker, R. D., Benbasat, I., Davis, F. D., Dennis, A. R., Gefen, D., Gupta, A., Ischebeck, A., Kenning, P., Müller-Putz, G., Pavlou, P. A., Riedl, R., vom Brocke, J., & Weber, B. (2012). On the use of neurophysiological tools in IS research: Developing a research agenda for NeuroIS. *MIS Quarterly*, *36*(3), 679-702.

Dimoka, A., Pavlou, P. A., & Davis, F. D. (2007). *NEURO-IS: The potential of cognitive neuroscience for information systems research*. Proceedings of the 28th International Conference on Information Systems, 1-20.

Dimoka, A., Pavlou, P. A., & Davis, F. D. (2011). NeuroIS: The potential of cognitive neuroscience for information systems research. *Information Systems Research*, *22*(4), 687-702.

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, *5*(2), 155-174.
Fabiani, M., Gratton, G., Karis, D., & Donchin, E. (1987). The definition, identification, and reliability of measurement of the P300 component of the event-related brain potential. In P. K. Ackles, J. R. Jennings, & M. G. H. Coles (Eds.), *Advances in psychophysiology* (pp.1-78). Greenwich, CT: JAI Press.
Fliessbach, K., Rohe, T., Linder, N. S., Trautner, P., Elger, C. E., & Weber, B. (2010). Retest reliability of reward-related BOLD signals. *Neuroimage*, *50*(3), 1168-76.
Foley, P., & Kirschbaum, C. (2010). Human hypothalamus-pituitary-adrenal axis responses to acute psychosocial stress in laboratory settings. *Neuroscience and Biobehavioral Reviews*, *35*(1), 91-96.
Fridlund, A. J., & Cacioppo, J. T. (1986). Guidelines for human electromyographic research. *Psychophysiology*, *23*(5), 567-89.
Gale, A., & Baker, S. (1981). In vivo or in vitro? Some effects of laboratory experiments, with particular reference to the psychophysiology experiment. In M. J. Christie, & P. G. Mellett (Eds.), *Foundations of psychosomatics*. Chichester: Wiley.
Gazzaniga, M. S., Russell, T., & Senior, C. (2009). *Methods in mind*. Cambridge: MIT Press.
Gratton, G. (2007). Biosignal processing. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (3rd edn.) (pp. 834-858). Cambridge: Cambridge University Press.
Guijt, A. M., Sluiter, J. K., & Frings-Dresen, M. H. W. (2007). Test-retest reliability of heart rate variability and respiration rate at rest and during light physical activity in normal subjects. *Archives of Medical Research*, *38*(1), 113-120.
Habermas, J. (1990). *Moral consciousness and communicative action*. Cambridge: Polity Press.
Hevner, A., Davis, C., Collins, R. W. & Gill, T. G. (2014). A NeuroDesign model for IS research. *Informing Science: The International Journal of an Emerging Transdiscipline*, *17*, 103-132.
Horvat-Gordon, M., Granger, D. A., Schwartz, E. B., Nelson, V. J., & Kivlighan, K. T. (2005). Oxytocin is not a valid biomarker when measured in saliva by immunoassay. *Physiology and Behavior*, 84(3), 445-448.
Huettel, S. A., Song, A. W. & McCarthy, G. (2009). *Functional magnetic resonance imaging* (2nd ed.). Massachusetts: Sinauer.
Jennings, J. R., & Gianaros, P. J. (2007). Methodology. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (3rd edn.) (pp. 812-833). Cambridge: Cambridge University Press.
Jin, P. (1992). Toward a reconceptualization of the law of initial value. *Psychological Bulletin*, *111*(1), 176-84.
Kantz, H., Kurths, J., & Mayer-Kress, G. (1998). *Nonlinear analysis of physiological data*. Berlin: Springer.
Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of behavioral research* (4th edn.). Wadsworth: Thomson.
Koch, G. G. (1982). Intraclass correlation coefficient. In S. Kotz & N. L. Johnson (Eds.), *Encyclopedia of statistical sciences 4* (pp. 213-217). New York: John Wiley & Sons.
Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, *28*, 563–575.
Léger, P.-M., Riedl, R., & vom Brocke, J. (2014). Emotions and ERP information sourcing: The moderating role of expertise. *Industrial Management & Data Systems*, *114*(3), 456-471.
Léger, P.-M., Sénecal, S., Courtemanche, F., Ortiz de Guinea, A., Titah, R., Fredette, M., Labonte-Lemoine, È. (2014). Precision is in the eye of the beholder: Application of eye fixation-related potentials to information systems research. *Journal of the Association for Information Systems*, *15*(10), 651-678.
Loos, P., Riedl, R., Müller-Putz, G. R., vom Brocke, J., Davis, F. D., Banker, R. D., & Léger, P.-M. (2010). NeuroIS: Neuroscientific approaches in the investigation and development of information systems. *Business & Information Systems Engineering*, *2*(6), 395-401.
Luczak, H., & Göbel, M. (2000). Signal processing and analysis in application. In R. W. Backs (Eds.), *Engineering psychophysiology: Issues and applications* (pp. 79-110). New Jersey: Lawrence Erlabaum Associates.
MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS Quarterly*, *35*(2), 293-334.
McCullough, M. E., Churchland, P. S., & Mendez, A. J. (2013). Problems with measuring peripheral oxytocin: Can the data on oxytocin and human behavior be trusted. *Neuroscience & Biobehavioral Reviews*, *37*(8), 1485-1492.

McEvoy, L. K., Smith, M. E., & Gevins, A. (2000). Test-retest reliability of cognitive EEG. *Clinical Neurophysiology*, *111*(3), 457-463.

Mingers, J. (2001). Combining IS research methods: Towards a pluralist methodology. *Information Systems Research*, *12*(3), 240-259.

Mukherjee, S., Yadav, R., Yung, I., Zajdel, D. P., & Oken, B. S. (2011). Sensitivity to mental effort and test-retest reliability of heart rate variability measures in healthy seniors. *Clinical Neurophysiology*, *122*(10), 2059-2066.

Müller, R., & Büttner, P. (1994). A critical discussion of intraclass correlation coefficients. *Statistic in Medicine*, *13*(23-24), 2465-2476.

Näpflin, M., Wildi, M., & Sarnthein, J. (2007). Test-retest reliability of resting EEG spectra validates a statistical signature of persons. *Clinical Neurophysiology*, *118*(11), 2519-2524.

Näpflin, M., Wildi, M., & Sarnthein, J. (2008).Test-retest reliability of EEG spectra during a working memory task. *Neuroimage*, *43*(4), 687-693.

O'Donnel, R. D., & Eggemeier, F. T. (1986). Workload assessment methodology. In Boff, K. R., Kaufman, L., & Thomas, J. P. (Eds.), *Handbook of perception and human performance—cognitive processes and performance* (pp. 42.1-42.49). New York: Wiley.

Ortiz de Guinea, A., & Webster, J. (2013). An investigation of information systems use patterns: Technological events as triggers, the effects of time, and consequences for performance. *MIS Quarterly*, *37*(4), 1165-1188.

Ortiz de Guinea, A., Titah, R., & Léger, P.-M. (2013). Measure for measure: A two-study multi-trait multi-method investigation of construct validity in information systems research. *Computers in Human Behavior*, *29*(3), 833-844.

Ortiz de Guinea, A., Titah, R., & Léger, P.-M. (2014). Explicit and implicit antecedents of users' behavioral beliefs in information systems: A neuropsychological investigation. *Journal of Management Information Systems*, *30*(4), 179-210.

Parasuraman, R., & Rizzo, M. (2008). Introduction to neuroergonomics. In R. Parasuraman & M. Rizzo (Eds.), *Neuroergonomics: The brain at work* (pp. 3-12). Oxford: Oxford University Press.

Pizzagalli, D. A. (2007). Electroencephalography and high-density electrophysiological source localization. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (3rd edn.) (pp. 56-84). Cambridge: Cambridge University Press.

Plichta, M. M., Schwarz, A. J., Grimm, O., Morgen, K., Mier, D., Haddad, L., Gerdes, A. B., Sauer, C., Tost, H., Esslinger, C., Colman, P., Wilson, F., Kirsch, P., & Meyer-Lindenberg, A. (2012). Test-retest reliability of evoked BOLD signals from a cognitive-emotive fMRI test battery. *Neuroimage*, *60*(3), 1746-1758.

Riedl, R. (2009). Zum Erkenntnispotenzial der kognitiven Neurowissenschaften für die Wirtschaftsinformatik: Überlegungen anhand exemplarischer Anwendungen. *NeuroPsychoEconomics*, *4*(1), 32-44.

Riedl, R. (2013). On the biology of technostress: Literature review and research agenda. *DATA BASE for Advances in Information Systems*, *44*(1), 18-55.

Riedl, R., & Javor, A. (2012). The biology of trust: Integrating evidence from genetics, endocrinology and functional brain imaging. *Journal of Neuroscience, Psychology, and Economics*, *5*(2), 63-91.

Riedl, R., Banker, R. D., Benbasat, I., Davis, F. D., Dennis, A. R., Dimoka, A., Gefen, D., Gupta, A., Ischebeck, A., Kenning, P., Müller-Putz, G., Pavlou, P. A., Straub, D. W., vom Brocke, J., & Weber, B. (2010a). On the foundations of NeuroIS: Reflections on the Gmunden Retreat 2009. *Communications of the Association for Information Systems*, *27*, 243-264.

Riedl, R., Hubert, M., & Kenning, P. (2010b). Are there neural gender differences in online trust? An fMRI study on the perceived trustworthiness of eBay offers. *MIS Quarterly*, *34*(2), 397-428.

Riedl, R., Kindermann, H., Auinger, A., & Javor, A. (2012). Technostress from a neurobiological perspective: System breakdown increases the stress hormone cortisol in computer users. *Business & Information Systems Engineering*, *4*(2), 61-69.

Riedl, R., Kindermann, H., Auinger, A., & Javor, A. (2013). Computer breakdown as a stress factor during task completion under time pressure: Identifying gender differences based on skin conductance. *Advances in Human-Computer Interaction*, *2013*(7).

Riedl, R., Mohr, P., Kenning, P., Davis, F., & Heekeren, H. (2014). Trusting humans and avatars: A brain imaging study based on evolution theory. *Journal of Management Information Systems*, *30*(4), 83-113.

Salinsky, M. C., Oken, B. S., & Morehead, L. (1991). Test-retest reliability In EEG frequency-analysis electroencephalography. *Clinical Neurophyisiology*, *79*(5), 382-392.

Schell, A. M., Dawson, M. E., Nuechterlein, K. H., Subotnik, K. L., & Ventura, J. (2002). The temporal stability of electrodermal variables over a one-year period in patients with recent-onset schizophrenia and normal subjects. *Psychophysiology*, *39*(2), 124-132.

Schultheiss, O. C., & Stanton, S. J. (2009). Assessment of salivary hormones. In E. Harmon-Jones & J. S. Beer (Eds.), *Methods in social neuroscience* (pp. 17-44). Guilford: New York.

Simon, H. A. (1980). The behavioral and social sciences. *Science*, *209*(4452), 72-78.

Straub, D. W. (1989). Validating instruments in MIS research. *MIS Quarterly*, *13*(2), 147-169.

Straub, D. W., Ang, S., & Evaristo, R. (1994). Normative standards for MIS research. *DATA BASE for Advances in Information Systems*, *25*(1), 21-34.

Straub, D. W., Gefen, D., & Boudreau, M.-C. (2004). Validation guidelines for IS positivist research. *Communications of the Association for Information Systems*, *13*(24), 380-427.

Stroobant, N., & Vingerhoets, G. (2001). Test-retest reliability of functional transcranial Doppler ultrasonography. *Ultrasound in Medicine & Biolology*, *27*(4), 509-14.

Strube, M. J., & Newman L. C. (2007). Psychometrics. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (3rd edn.) (pp. 789-811). Cambridge: Cambridge University Press.

Tams, S., Hill, K., Ortiz de Guinea, A., Thatcher, J., & Grover, V. (2014). NeuroIS—alternative or complement to existing methods? Illustrating the holistic effects of neuroscience and self-reported data in the context of technostress research. *Journal of the Association for Information Systems*, *15*(10), 723-752.

Tervaniemi, M., Lehtokoski, A., Sinkkonen, J., Virtanen, J., Ilmoniemi, R. J., & Näätänen R. (1999). Test-retest reliability of mismatch negativity for duration, frequency and intensity changes. *Clinical Neurophysiology*, *110*(8), 1388-1393.

Thesen, T., & Murphy, C. (2002). Reliability analysis of event-related brain potentials to olfactory stimuli. *Psychophysiology*, *39*(6), 733-738.

Vance, A., Anderson, B., Kirwan, C. B., & Eargle, D. W. (2014). Using measures of risk perception to predict information security behavior: Insights from electroencephalography (EEG). *Journal of the Association for Information Systems*, *15*(10), 679-722.

Venkatesh, V., Brown, S. A., & Bala, H. (2013). Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems. *MIS Quarterly*, *37*(1), 21-54.

Vogel, D. R., & Wetherbe, J. C. (1984). MIS research: A profile of leading journals and universities. *DATA BASE for Advances in Information Systems*, *16*(1), 3-14.

vom Brocke, J., & Liang, T.-P. (2014). Guidelines for neuroscience studies in information systems research. *Journal of Management Information Systems*, *30*(4), 211-234.

vom Brocke, J., Riedl, R., & Léger, P.-M. (2013). Application strategies for neuroscience in information systems design science research. *Journal of Computer Information Systems*, *53*(3), 1-13.

Weil, M. M., & Rosen, L. D. (1997). *Technostress: Coping with technology work home play*. New York: Wiley.

Wilder, J. (1967). *Stimulus and response: The law of initial value*. Bristol, UK: John Wright & Sons.

About the Authors

René RIEDL is a professor for Digital Business and Innovation at the University of Applied Sciences Upper Austria and an associate professor for Business Informatics at the University of Linz. Moreover, he serves on the executive board of the Institute of Human Resources and Organizational Development in Management (IPO) at the University of Linz. He has published in the following outlets, among others: *Advances in Human-Computer Interaction*, *Behavior Research Methods*, *Business & Information Systems Engine*ering, *Communications of the AIS*, *DATA BASE for Advances in Information Systems*, *Industrial Management & Data Systems*, *Journal of Computer Information Systems*, *Journal of Information Technology*, *Journal of Management Information Systems*, *Journal of Neuroscience*, *Psychology, and Economics*, and *MIS Quarterly*. He holds or has held various editorial positions (*AIS Transactions on Human-Computer Interaction*, B*usiness & Information Systems Engineering*, *DATA BASE for Advances in Information Systems*, *Information Systems Journal*, *Journal of Information Theory and Applica*tion, *Journal of Management Information Systems*, *Journal of the AIS*, and *MIS Quarterly*).

Fred D. DAVIS is Distinguished Professor and David D. Glass Chair in Information Systems at the Sam M. Walton College of Business, University of Arkansas. He served as Chair of the Information Systems Department at the Walton College from 1999-2007. After earning his Ph.D. from MIT Sloan School of Management, he served on the business school faculties of the Universities of Michigan, Minnesota, and Maryland. He teaches classes at the PhD, Masters, and undergraduate level on topics including research methodology, business analytics, business intelligence, decision support systems, business problem solving, and statistical analysis. He is a Senior Editor for *Information Systems Research* and the *AIS Transactions on Human-Computer Interaction*, and was Associate Editor for *Management Science*, *MIS Quarterly*, and *Information Systems Research*. His research interests include user acceptance of information technology, technology supported decision making, computer training and skill acquisition, NeuroIS, and system development practices. His research has been published in such journals as *MIS Quarterly*, *Information Systems Research*, *Management Science*, *Journal of MIS*, *Journal of the AIS*, *Decision Sciences*, *Journal of Applied Psychology*, *Journal of Experimental Social Psychology*, *Journal of Applied Social Psychology*, *Organizational Behavior and Human Decision Processes*, *IEEE Transactions*, and *International Journal of Human-Computer Studies*.

Alan R. HEVNER (PhD, Computer Science, Purdue University) is an Eminent Scholar and Professor in the Information Systems and Decision Sciences Department in the College of Business at the University of South Florida. He holds the Citigroup/Hidden River Chair of Distributed Technology. Dr. Hevner's areas of research interest include design science research, information systems development, software engineering, distributed database systems, healthcare systems, and service-oriented computing. As a thought leader on design science research in information systems, Dr. Hevner has co-authored a 2010 book on design science research, presented seminars internationally on the topic, and co-founded an international conference (Design Science Research in Information Systems and Technology—DESRIST). He has published over 200 research papers on these topics and has consulted for a number of Fortune 500 companies. He has held faculty positions at the University of Maryland College Park and the University of Minnesota. He is a Fellow of the American Association for the Advancement of Science (AAAS) and is a member of ACM, IEEE, AIS, and INFORMS. From 2006 to 2009, he served as a program manager at the U.S. National Science Foundation (NSF) in the Computer and Information Science and Engineering (CISE) Directorate.