Expanding the Criteria for Evaluating Socio-Technical Software

Brian Whitworth, Victor Bañuls, Cheickna Sylla, and Edward Mahinda

Abstract—**This paper compares two evaluation criterion frameworks for sociotechnical software. Research on the technology acceptance model (TAM) confirms that perceived usefulness and perceived ease of use are relevant criteria for users evaluating organizational software. However, information technology has changed considerably since TAM's 1989 inception, so an upgraded evaluation framework may apply. The web of system performance (WOSP) model suggests eight evaluation criteria, based on a systems theory definition of performance. This paper compares WOSP and TAM criterion frameworks in a performance evaluation experiment using the analytic hierarchy process method. Subjects who used both TAM and WOSP criteria preferred the WOSP criteria, were more satisfied with its decision outcomes, and found the WOSP evaluation more accurate and complete. As sociotechnical software becomes more complex, users may need (or prefer) more comprehensive evaluation criterion frameworks.**

Index Terms—**Sociotechnical, software requirements, system performance, technology assessment.**

I. Introduction

OVER the last two decades, over 50% of new organizational capital investments have been in information technology (IT) [1]. The 2001 total worldwide IT expenditure exceeded one trillion U.S. dollars per annum, with an expected annual compounded growth rate of 10% [2]. Modern businesses must decide whether to purchase (or upgrade to) emergent technology in various states of maturity. Purchasing new but immature IT can cost a business as much as failing to upgrade to the next wave of technology advances. Given both tangible risks and intangible opportunities, evaluating new technology has become a critical business survival need. New technology evaluation is a multibillion-dollar issue that affects all organizations that use IT.

Organizations can get more value from expensive IT by better new product evaluation, i.e., by "buying smarter." This enhances overall performance [3] and gives executives the information needed to justify huge IT investments [4]. Despite automated technology evaluation research [5], most technology evaluation still involves people, perhaps because of the following reasons.

1) People understand the organizational context within which IT evaluations occur.
2) People understand the IT products that users prefer.

Experienced staff with tacit knowledge can evaluate new software, as they know both what the organization wants and what end users want. How then do evaluations proceed? To make a rational decision, which can be justified to others, requires explicit criteria, by which one outcome is compared with another. An evaluation without criteria has no cognitive structure [6] to argue that one outcome is "better" than another. Rational decision making requires valid criteria [7], and changing criteria can make a problem a "moving target" [8]. This paper investigates the criteria of technology evaluation by comparing two selected frameworks. Section II reviews current technology evaluation frameworks, Section III introduces a new criterion framework, Section IV presents a strategy to compare criterion frameworks, Section V gives the method, Section VI analyzes the results, and Section VII discusses limitations and implications. The criteria and statements used are given in Appendix I.

II. Current Technology Evaluation Criteria

A. Perspectives

Three perspectives seem to permeate the study of people and technology, each with a different research culture, different journals, and different conferences, which are summarized as follows [9].

1) Human factors (ergonomics) research: the perspective of system designers who add user needs to existing requirements. How to code the system is presumed known, but what the user wants is not, giving the question: "What do users want from software?"
2) Computer–human interaction research: the perspective of a user generation who can develop systems to fit their own needs. Now, the user (myself) is known, but how to get a computer to satisfy my needs is not, giving the question: "How can I get software to do what I want?"
3) Management information systems research: the perspective of managers who want their staff to accept purchased technology, giving the question: "What makes users accept new software?"

Each perspective takes the viewpoint of a stakeholder in technology creation, namely, that of the system designer, the

Manuscript received November 25, 2005; revised January 14, 2007. This work was supported in part by grants from the Fulbright Foundation, Washington, D.C., and by NJIT Research Office. This paper was recommended by Associate Editor L. Rothrock.

B. Whitworth is with the Institute of Information and Mathematical Sciences, Massey University, Albany, Auckland 0745, New Zealand (e-mail: bwhitworth@acm.org).

V. Bañuls is with the Management Department, Pablo de Olavide University, 41013 Seville, Spain (e-mail: vabansil@upo.es).

C. Sylla is with the School of Management, New Jersey Institute of Technology, Newark, NJ 07102-1982 USA (e-mail: sylla@adm.njit.edu).

E. Mahinda is with the Department of Information Systems, New Jersey Institute of Technology, Newark, NJ 07102-1982 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMCA.2008.923038

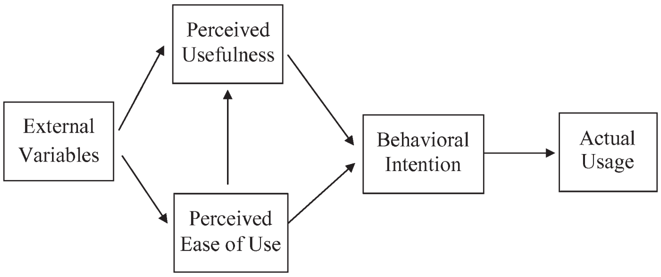

Fig. 1. Technology acceptance model.

end user, and the IT manager. The fact that introducing new technology involves designers, users, and managers suggests that its evaluation criteria should be meaningful to all, as the following criterion scenarios suggest.

1) Criteria useful to users only, like "attractiveness," let users know when they see it, but designers have difficulty designing it, and managers find it difficult to request.
2) Criteria useful to designers only, like object-oriented programming, allow designers to implement it, but managers and users struggle to fund and relate to it.
3) Criteria useful to managers only, like infrastructure compatibility, help managers but are difficult for designers to predict, and for users, they are a background context.

It seems desirable that IT evaluation criteria be relevant to all IT stakeholders, i.e., to the designers who create IT, to the users who use IT, and to the managers who fund IT.

B. Technology Acceptance Model (TAM)

The TAM derives from the theory of reasoned action [10] and is shown in Fig. 1 [11]. TAM implied that what a system does is an incomplete description of its performance and added a new technology acceptance criterion called perceived ease of use (PEOU) to the established perceived usefulness (PU) [12]. TAM made performance bidimensional, with usability and usefulness being relatively distinct criteria [13]. This then explained how functional but unfriendly Web systems "failed"—their users found them difficult and simply "clicked on" to more friendly Web sites.

While PU and PEOU are user perceptions, they are "grounded" in system stimuli, i.e., systems with more functions are generally seen as more useful, and systems with better interfaces are generally seen as easier to use. The cognitive criteria PU and PEOU can map to the design criteria of functionality and usability. TAM is thus a theory of user perceptions of design criteria, making it relevant to designers, users, and managers. This relevance to all IT creation stakeholders could explain its extraordinary success.

However, while studies have validated TAM's constructs in general [11], for Web sites [14], for online shopping [15], for Internet banking [16] and for Web portals [17], other work suggests that TAM is incomplete. As a TAM review notes: "...even if established versions include additional variables, the model hardly explains more than 40% of the variance in use." [17, p. 202]. Recognition of TAM's incompleteness has increased over the years since its inception, leading to many suggested extensions.

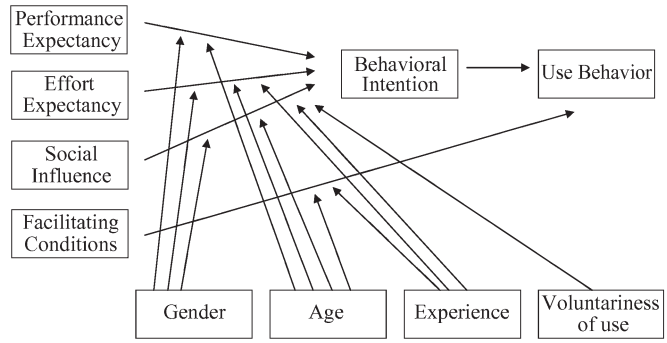

Fig. 2. UTAUT model [28].

C. Psychological Extensions

TAM's psychological extensions include variables that moderate PU and PEOU, like gender [19], experience [3], and culture [20], and PU and PEOU antecedents, like self-efficacy, external control, computer anxiety, and playfulness [21]. In general, researchers sought to extend TAM using cognitive variables that predicted acceptance also, like playfulness [22], credibility [23], attractiveness [17], self-efficacy [23], behavioral control [3], user satisfaction [24], and enjoyment [25]. Yet, how these variables relate to PU and PEOU, and to each other, has led to "...a state of theoretical chaos and confusion in which it is not clear which version of the many versions of TAM is the commonly accepted one." [26, p. 2].

One problem with psychological criteria is that perceptions gathered from the same subjects at the same time can confound because they originate in the same mind, e.g., computer anxiety may reduce ease of use, but hard to use software may also cause anxiety. In cognitive dissonance theory, people mutually adjust their internal attitudes to be consistent [27], so cognitive correlations need not imply causal relations. Even PU and PEOU, while conceptually distinct, correlate significantly, and each "predicts" the other [20], [23]. Correlated technology acceptance cognitions may reflect more about how people think than about technology acceptance. While TAM's original PU and PEOU are relevant to users, designers, and managers, most proposed cognitive add-ons are not, e.g., what is the design equivalent of enjoyment or self-efficacy? How do managers buy software that is "credible?" While psychology adds useful moderating and antecedent variables, adding psychological depth to TAM does not add breadth—it is still a 2 factor performance model.

D. Organizational Extensions

Organizational extensions to TAM concern the business context within which the software operates. A recent unified theory of acceptance and use of technology (UTAUT) model (Fig. 2) combined eight previous psychological and sociological models with TAM [28]. UTAUT renamed TAM's original variables to *performance expectancy* and *effort expectancy*, then added psychology moderators like gender, and finally added the following two well-known organizational constructs.

1) Social influence: The degree users believe that important others support the system.
2) Facilitating technology: The degree users believe that organizational and technical infrastructure exists to support the system.

Social influence suggests that users will accept new technology when significant others have already done so. Such "conformity" has a long history in social research [29]–[33]. Facilitating technology suggests that organizations select software that fits their existing infrastructure, again a well-known effect. These two "inertial" factors describe social and practical barriers to innovation, but they do not explain how new technology takes hold in the first place. New products, like bluetooth, by definition, often begin without either infrastructure or normative support. What distinguishes products like cell phones, which take off, from those that do not, like the video phone? If UTAUT's inertial factors oppose most innovations, then, for new technology evaluations, UTAUT's predictive variables effectively reduce to performance and effort expectancy, i.e., collapse to TAM's original variables. UTAUT suggests that, if a new product is useful and usable, marketing and infrastructure support will guarantee success. "Mr. Clippy" (Microsoft's Office Assistant) was Bayesian smart, user friendly, and well marketed and supported, so both TAM and UTAUT predicted his success. Yet, he was so notable a failure that Mr. Clippy's removal was a Windows XP promotion pitch [34].

For new technology evaluation, UTAUT offers little beyond what TAM offers already. What is needed is not a collage of theories from different fields, but a return to TAM's original direction, of constructs that cut across stakeholders.

E. System Levels

Technology evaluation suggests the following three variable types [35]:

1) System variables: Is it useful, easy to use, secure, etc.?
2) User variables: Age, gender, experience, attitude, etc.
3) Organizational variables: Corporate goals, technology infrastructure, social structures, normative influences, etc.

Each variable type has its own academic specialty. System factors such as security exist in system engineering models [36]. User variables such as computer anxiety exist in psychology models, like social cognitive theory [37]. Organizational characteristics such as normative effects exist in sociology models, like innovation diffusion theory [38]. Yet, these models are all relevant to IT evaluation. One way to connect them conceptually is to postulate IS "levels" to match the previous variable types [39]. Grudin [40] has suggested three levels: hardware, software, and cognitive, whereas Kuutti [41] adds a social work processes level. We also suggest the following four system levels:

level 1) hardware: computers, wires, printer, keyboard, and mouse;
level 2) software: programs, data, bandwidth, and memory;
level 3) human–computer interaction (HCI): attitudes, beliefs, opinions, and ideas;
level 4) sociotechnical: norms, culture, laws, sanctions, and roles.

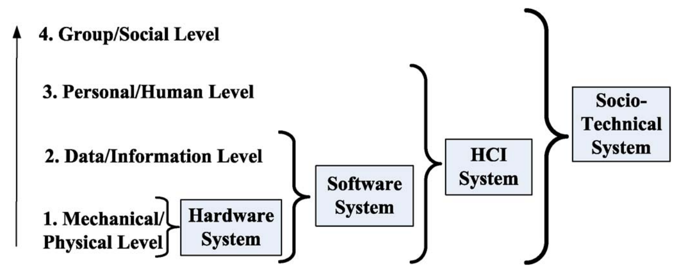

Fig. 3. IS/IT levels and systems.

These four levels match the disciplines of engineering, computing, psychology, and social sciences, respectively. They also match Alter's view of information systems (IS) as hardware, software, people, and processes [42]. They combine to create four system types (see Fig. 3). Hardware and software gives technology. People plus technology gives HCI systems (hardware + software + people), which can exchange meaning as well as data or information (where "information" is as Shannon and Weaver [43] define it). Finally, when individuals form communities mediated by technology, the result is a four-level sociotechnical system (STS) [44]. IS users, indeed, seem to exchange factual, personal, and group information using different cognitive processes, corresponding to levels 2, 3, and 4 earlier [33], [45].

In this view, an STS is a social system built upon a technical system. One could equally consider normal social interaction as a "sociophysical system" (a social system built upon a physical one). Each IS/IT level "emerges" from the previous and is a prerequisite for the next, i.e., physical actions create data, data create human meaning, and many human meanings combine into a group identity. Higher levels are not new systems but higher "views" of the same system. While a computer's hardware is fully described by circuits and voltages, it can, at the same time, be fully described by software data flow diagrams, etc. Yet, there is only one system. Hence, "hardware" and "software" are simply different ways of seeing the same thing. Likewise, HCI and sociotechnical levels merely add personal and social perspectives to technology systems.

Higher levels are not only more efficient ways to view a system but also better ways to operate it, e.g., better protocols can improve network performance just as better cables can. However, higher level benefits have corresponding higher level requirements, e.g., after hardware requirements like overheating are met, software requirements like information throughput and data record locking arise to drive system design and architecture.

A critical feature of TAM is that it connects higher and lower IS levels—it links user requirements (HCI level) and technology design. TAM's constructs are psychological, but they map to technical requirements. In contrast, most psychological and organizational extensions to TAM add variables with no system correlates, thus: ". . .repeatedly demonstrating that certain mediators (beliefs) are influential without understanding how to influence such beliefs through IT design is ultimately of limited value." [26, p. 7].

The conclusion that useful IS/IT models connect IS/IT levels generalizes the earlier conclusion that useful criteria impact all stakeholders. The logic is that user acceptance requires an information system and that information systems need system design, and therefore, system design is relevant to user acceptance. Based on this thinking, some have advocated a critical need to link IS research and system design by articulating IS theories with design consequences [57]. They argue that, if IS research does not contribute to IS design [58], it may become increasingly irrelevant to technology progress. This again suggests that TAM is best extended at its system design roots, with constructs like security, compatibility, and privacy [15].

F. System Engineering Criteria

Unfortunately, the systems engineering literature is not very clear about the criteria that define technology "performance." A recent software engineering text suggests the criteria of usability, repairability, security, and reliability [36, p. 24]. However, the ISO 9126-1 quality model suggests functionality, usability, reliability, efficiency, maintainability, and portability as critical criteria [46]. Berners-Lee [47], however, found scalability as the key to World Wide Web success, whereas others espouse open standards [48]. Alter [42] suggests the criteria of cost, quality, reliability, responsiveness, and conformance to standards. Software architects argue for portability, modifiability, and extendibility [49], whereas others find flexibility as a critical success factor [50]. Still, others suggest that privacy is what users really want [51]. On the issue of what criteria must human–machine systems satisfy to perform well, the system design literature is at best confused.

Even worse, performance categories are confounded, e.g., "security" by one definition is an umbrella term that includes availability, confidentiality, and integrity [53]. This makes reliability an aspect of security. However, "dependability" has been defined as reliability, availability, safety, and security [52], making security part of a larger reliability-type construct. Is reliability part of security or vice versa? Because each design specialty sees others as subsets of itself, reliability falls under security in security models, but security falls under reliability in reliability models. Yet, research shows that increasing fault tolerance (reliability) can reduce security and also that increasing security can cause breakdowns [54]. This suggests that neither category subsumes the other. Academic parochialism means that, while an ISO 9241-10 usability inventory measure finds "suitability for the task" (functionality) and "error tolerance" (reliability) aspects of a broad "usability" construct [55], another review finds "scalability," "robustness," and "connectivity" aspects of an equally general "flexibility" concept [50, p. 6]. Such competition for theoretical space creates confusion, not consensus, as each specialty tries to expand itself at the other's expense. Our approach is to return to the original goal of requirements engineering, namely: "The primary measure of success of a software system is the degree to which it meets the purpose for which it was intended. Broadly speaking, software systems requirements engineering (RE) is the process of discovering that purpose. . ." [56].

This purpose, it is proposed, is system performance. Yet, if designers, users, and managers all have different purposes, how can a common "performance" be defined? One way is to take the view of the system itself, which we now do.

III. General Systems Approach

The web of system performance (WOSP) model uses a general systems perspective [59] to define performance and decompose it into a multigoal model as suggested by Chung [60]. A brief description follows, as a detailed justification is given elsewhere [61], [62]. The model assumes only a general "system," so it can apply at both technical and social levels. WOSP is less a design theory than a theory about the nature of system design, from which design theories can be derived. While the WOSP model's origins lie in design and those of TAM in acceptance, both imply evaluation criteria.

A. WOSP

A system is an entity within a "world," whose nature defines the nature of the system. The world need not be physical, so if information, cognitive, and social "worlds" exist, then information systems, cognitive systems, and social systems, respectively, can exist within them. A system's "environment" is that part of the world that affects the system, for benefit or harm. The WOSP model takes system performance as how well the system survives and prospers in its environment, given that what succeeds in one environment may fail in another. The WOSP model suggests the following four system elements.

1) *Boundary* separates the system from the environment.
2) *Internal structure* supports and coordinates the system.
3) *Effectors* act upon the environment.
4) *Receptors* analyzes environment information.

For example, people have a skin boundary, internal brain and organs, acting muscles, and eyes and ears as sensory receptors. Computers have a physical case boundary, an internal architecture, printer/screen effectors, and keyboard and mouse "receptors."

Four system elements by two environment outcomes (gain and loss) gives eight performance goals (each element can increase gains or reduce losses). The *boundary* controls system entry, so it can be designed to deny an unwelcome entity entry (security) or to use the entity as a "tool" (extendibility). The *internal structure* manages a system's operations, so it can be designed to reduce the internal changes that cause faults (reliability) or to increase the internal changes that allow adaptation to environment changes (flexibility). The *effectors* use system resources to act upon the environment, so it can be designed to maximize their effects (functionality) or to minimize the relative resource "cost of action" (usability). Finally, *receptors* open channels to communicate with the world, so systems can enable communication (connectivity) or limit it (privacy). None of the eight performance goals of Table I is new, as similar concepts pervade the IS/IT literature, but this model integrates

TABLE I
SYSTEM PERFORMANCE GOALS

System Element	Goal	Definition
Boundary	*Security*	To protect against unauthorized entry, misuse or takeover
	Extendibility	To use outside elements as part of the system
Internal structure	*Flexibility*	To adapt the system's operation to environment changes
	Reliability	To continue operating despite internal failure
Effector	*Functionality*	To act directly on the environment to produce a desired change
	Usability	To minimize the relative resource costs of action
Receptor	*Connectivity*	To open and use channels to communicate with other systems
	Privacy	To manage the release of self information

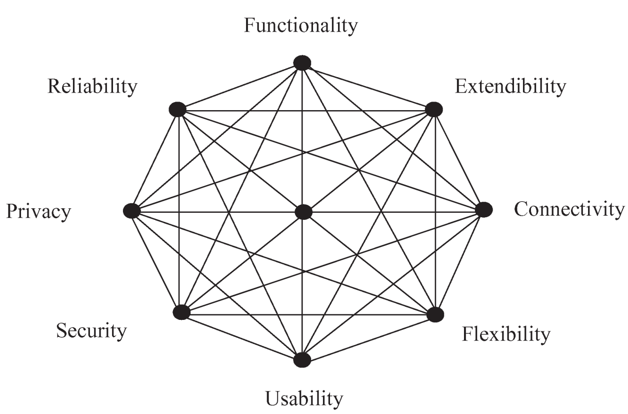

Fig. 4. Web of system performance.

them into a common framework (Fig. 4), where the following are observed.

1) The web area is the overall system performance. A larger area has more performance potential.
2) The web shape shows each goal's contribution to performance in a given environment, e.g., turbulent environments favor reliability.
3) The web lines show goal tensions, imagined as rubber connecting bands, e.g., improving flexibility might reduce reliability.

B. WOSP Criteria

The WOSP performance criteria have the following features.

1) Criterion validity: The criteria derive from the general nature of systems and are not assumed as self-evident.
2) Criterion equivalence: Any or all criteria can be critical, and different environments weight them differently.
3) Criterion modularity: The criteria are logically distinct.
4) Criterion tension: Designing to satisfy one criterion can reduce one or more others.
5) Multilevel applicability: Criteria can apply at any IS level (but only to one level at a time).

1) Criteria Validity: Unlike models that simply state relevant constructs, the WOSP model defines its general concern (performance) and then derives criteria. The results match well with known criteria terms, e.g., extendibility matches with openness, interoperability, permeability, compatibility, and scalability [47]. Flexibility matches adaptability, agility, portability, customizability, plasticity, and modifiability [63]. Reliability fits terms like stability, dependability, durability maintainability, and ruggedness. Finally, functionality and usability can connect to TAM's PU and PEOU. None of these correspondences was inevitable from the initial model.

2) Criteria Equivalence: Systems engineering often considers criteria like usability or reliability as "quality" requirements, which modify the primary functional goal, but cannot not "stand alone" [65]. For decades, these "-ilities" have stood apart from the main "what the system does" specification. They have defied categorization, as while most agree that they are important, there is little agreement on how they relate to performance (or each other). The WOSP model queries the assumption that "nonfunctional" requirements (NFRs) differ essentially from functional ones [66], as it does not distinguish "functional" and "nonfunctional" performance goals.

Consider the argument that NFRs like reliability are secondary because they cannot exist without functionality. This is true, but functionality also depends on reliability, as a system that cannot operate cannot function as well. The same logic that makes reliability secondary to functionality also makes functionality secondary to reliability. It is not easy to argue that functionality is "obviously" more important, as many software failures involve NFRs [67, p. 699], and the error modules that create reliability often involve more codes than the functional mainline (as do the interface modules that create usability).

The WOSP perspective is that functionality is simply the criterion that was historically most evident to a user, e.g., connectivity today seems as critical to IS/IT performance as functionality. WOSP has four risk-reducing criteria (security, reliability, privacy, and usability) and four opportunity-enhancing criteria (functionality, extendibility, flexibility, and connectivity). Because reducing risk is as important to system performance as increasing success [68], no distinction is made between "active" and "passive" WOSP goals. If not satisfying a performance criteria, like usability, can cause system failure, then that criterion defines as well as modifies performance.

3) Criteria Modularity: Criteria modularity means that the criterion goals do not overlap or contradict in principle. As already noted, the goals of Table I interact in design tensions, but criteria modularity means that they do not inherently contradict. If achieving one criteria need not necessarily prevent another, all combinations are possible, e.g., a bulletproof plexiglass room illustrates high security with no privacy at all, and perfect encryption lets one lose data (insecurity) but retain privacy. In the WOSP model, reliability and security are distinct performance goals, with the first aiming to provide services and the second to deny services [54]. Hence, a system can be reliable and secure, reliable and insecure, unreliable and secure, or unreliable and insecure. All WOSP dimensions are modular in the same way. Therefore, functionality is not the inevitable enemy of usability [70], nor does connectivity mean the end of privacy [71].

4) Criterion Tension: WOSP supports the view that reliability and flexibility are internal and external aspects of a common "robustness" concept in a natural design tension [64]. In the system requirements literature, such interacting goals are "cross-cutting requirements" [72], for example:

> "The face of security is changing. In the past, systems were often grouped into ... those that placed security above all other requirements, and those for which security was not a significant concern. But ... pressures ... have forced even the builders of the most security-critical systems to consider security as only one of the many goals that they must achieve" [69].

Hence, designers must satisfy not only the eight Table I criteria but also their interactions. Fortunately, apparent opposites like flexibility and reliability are goals in tension, not simple tradeoffs, and can be "reconciled" by innovative design.

5) Multilevel Applicability: The WOSP model can apply to any IS level, e.g., a system can be hardware reliable but software unreliable or can be both hardware and software reliable but operator unreliable [36, p. 24]. The term "reliability" means different things at different IS levels. Likewise, "usability," on a personal level, means less cognitive "effort" but, on a software level, means less memory/processing (e.g., "light" background utilities) and, on a hardware level, means less power usage (e.g., mobile phones that last longer).

Different IS system levels are not comparable, as, to compare entities, one needs a consistent viewpoint, but by definition, each level is a different world view. Hence, theories must first choose a level and then define variables from that perspective. To do otherwise is to fail to recognize that a theory is itself also a point of view, e.g., while psychology and engineering theories show a system from different angles, combining them into a "metatheory" simply confuses. The WOSP goals change their nature at each IS level, so the system level must be defined before the model can be applied.

IV. Research Strategy

The research context of people evaluating software for an organization is sociotechnical, which suits the WOSP model, with its social dimension (connectivity–privacy). While individuals evaluating technology for organizations is not the setting for which TAM was originally devised, its criteria can be, and have been, applied to this case [35]. The IS level evaluated is software, but the evaluation itself is at the HCI level, as perceptions are gathered. For individuals evaluating corporate software, the relevant criteria are at the human–machine level, not the organizational level of other studies [73]. The WOSP model suggests the following criteria: perceived security, extendibility, flexibility, reliability, functionality, usability, connectivity, and privacy, whereas TAM suggests PU and PEOU. The research goal is to compare these two evaluation criterion frameworks in practical use.

A. Research Question

To compare WOSP and TAM criteria, their theoretical relation must be clear. If usefulness (or performance expectancy) is how the user believes the system will help his or her job performance, it could connect to WOSP's similar but more restricted functionality, a system's ability to act upon its environment. Ease of use (effort expectancy) could then map to WOSP's usability (cost of action). In this case, the remaining six WOSP factors would add new criteria to the TAM/UTAUT core constructs.

However, if TAM's usefulness maps to system performance in general, "usefulness" is now a bigger concept than WOSP's functionality, e.g., insecure systems impact job performance, so users may rate insecure systems as less "useful." In this case, WOSP still extends TAM by better specifying its now very general "usefulness" construct. Otherwise, research showing that "usefulness is useful" is like finding better performance is better, which may add little value [74]. However, this second case creates an inconsistency—if NFRs like security fall under a generic "usefulness" construct, why does the same logic not apply to ease of use? Is ease of use not useful? The WOSP model avoids this problem by putting usability under a generic performance construct along with the other factors.

Whether WOSP adds to TAM or better specifies TAM, it offers a different evaluation framework. While most TAM research investigates possible new variables, this paper investigates a possible new criterion framework. The research question is: Do users prefer WOSP to TAM criteria when evaluating software for organizational use?

B. Hypotheses

Just as one cannot know, in an absolute sense, that one software application is "better" than another, so one cannot say absolutely that one criterion framework is better than another. Yet, research merely aims to subject theories to falsification, not to prove them. One falsifiable expectation of the view that WOSP offers a better criterion framework is that it will give the following different software evaluations.

H1) The WOSP application evaluation rankings will differ from the TAM evaluation rankings.

Because the WOSP model specifies criteria not explicitly named by TAM, like privacy, a second falsifiable expectation is that criteria other than functionality and usability will contribute significantly to the evaluation.

H2) In the WOSP evaluation, other criteria will contribute equally or more than functionality and usability.

If the WOSP evaluation framework adds value, subjects should prefer it after trying both methods, even before seeing the evaluation results.

H3) Users prefer to use the WOSP rather than the TAM criteria to evaluate software for an organization.

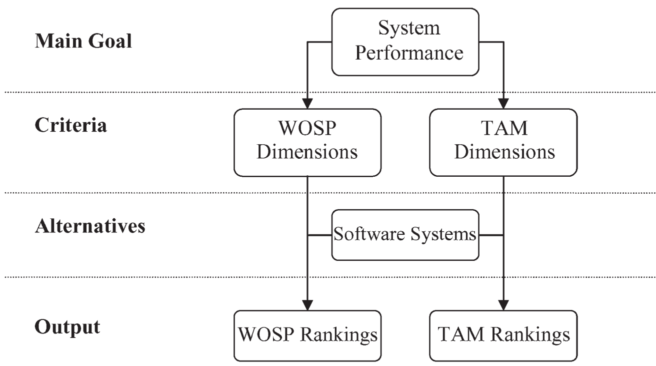

Fig. 5. AHP design.

Finally, users should be more satisfied with the evaluation outcome, feel more confident about it, and find the WOSP evaluation more accurate and complete.

H4) Users evaluating software for an organization will feel the following:
- H4a) more satisfied with the WOSP evaluation outcome;
- H4b) more confident in their choice using WOSP;
- H4c) that the WOSP evaluation was more accurate;
- H4d) that the WOSP evaluation was more complete.

C. Analytic Hierarchy Process (AHP) Tool

We used a criterion-based method of system evaluation to compare the WOSP and TAM frameworks. The AHP method [75], [76] is used for complex decision problems involving multiple criteria. The decision problem is formulated as a hierarchical structure, with the main goal at the top. Below that are the criteria and subcriteria used to judge the decision alternatives, which are below that again (Fig. 5). Pairwise comparison using decision-maker judgments is used to develop weights for the criteria, subcriteria, and alternatives (relative to the main goal). The AHP method is validated by the consistency of its decision matrix, as decision makers can be inconsistent (e.g., grading $a > b$, $b > c$, and $c > a$ is an inconsistent judgement). To detect inconsistencies, AHP uses the decision-matrix consistence ratio (CR), with an acceptable CR being less than 0.1 [76]. Although the AHP method has been applied to software selection previously [16], [71]–[73], we found no previous application of the TAM or WOSP models. Using the common goal of evaluating system performance, we transformed both models into a multicriterion decision hierarchy (Fig. 4) that allowed AHP comparisons.

V. Method

The research strategy was to let subjects use the two frameworks to rank software applications and then to measure the differences. We used a two-way randomized repeated-measure block factorial design, where each subject evaluated applications using both WOSP and TAM frameworks, with method and application orders controlled for. Subjects were told to try both evaluation methods and then to honestly indicate which they preferred. The study used behavioral intention to use, rather than actual use, as prior studies strongly link intention and behavior [35].

A. Software Evaluated

The modern Web browser is a powerful feature-rich application, which is increasingly important in personal and business use. It is the key software interface for hyperlinked resources on the Internet, as well as corporate intranets and extranets. Browsers are a universal software platform from which users can launch information searches, e-mail, multimedia file transfer, discussion groups, help files, and many other uses. We chose browsers as the evaluated software because they are commonly used, important in business, and have a social dimension. The browsers evaluated were as follows:

1) Internet Explorer (current version);
2) Mozilla Firefox;
3) MyIE2.

B. Task

Subjects were told that they were part of a large international company that needed a common Internet browser to minimize problems, standardize updates, and improve performance. Their task was to select the best performing browser for the company, from the short list provided.

C. Procedure

An external researcher completed the AHP procedure asynchronously, as it was impossible to get subjects together at the same time and place. Normally, AHP assessors meet face to face, but in this paper, all communication was by e-mail. Assessing eight WOSP dimensions across three applications seemed about the limit of subject capability, so this procedure could not be used, for example, to assess 30 browsers. The process involved the following four phases.

- Phase 1) Introduction: Subjects agreed to participate, signed the consent, answered a simple questionnaire, and were asked not to discuss the study with others.
- Phase 2) Criterion understanding: Subjects were given the criterion definitions and explanatory statements (Appendix I) and asked to rate them. This phase helped subjects understand the different criteria.
- Phase 3) AHP evaluation: The AHP method was explained, and subjects evaluated both browsers (TAM then WOSP or WOSP then TAM). For inconsistent responses, a reply explained the inconsistency, and subjects had to resubmit until consistent responses were achieved. After using both methods, subjects were asked if which they preferred.
- Phase 4) Outcome evaluation: The AHP responses were processed, and the results were given to each subject, who then completed a questionnaire on each method's outcome.

D. Criterion Understanding

Phase 2) helped subjects understand the criteria, as they had to not only read them but also to respond to them. No subjects queried either the TAM or WOSP definitions. In this "priming" phase, subjects were given the criteria and then had to respond on a 1–5 Likert scale (strongly disagree to strongly agree) if each statement was as follows:

1) Clear? Do you understand what it means?
2) Valid? Is it a valid statement of the definition given?
3) Important? Is it important when selecting a browser?

E. Subjects

The subjects were 28 students enrolled in an undergraduate management information systems evening course, with 65% males and 35% females. They were a culturally diverse group, and most also worked full time. Subjects, on average, had used browsers for 7.75 years and spent over 26 h/week using them, so were experienced browser users.

Each participant had four days to complete each of the four phases of the experiment, which ran over three weeks. If a phase response was not received within four days, the subject was eliminated from the next experiment phase. Phase 2) subjects made 41 AHP comparisons and repeated some if they were inconsistent. Phases 2) and 3) took 1–2 h to complete or longer. One subject was rejected in the second phase, five in the third phase, and two in the fourth phase. The initial 28 subjects became 20, as 8 were eliminated by time and quality requirements. Such dropouts are normal in AHP's lengthy and rigorous process, but the benefit is that the remaining subjects are strongly committed.

F. Variables

The experiment independent variable was the evaluation criterion framework (TAM or WOSP). The dependent variables were the application rankings, criterion weights, method preference, and outcome attitudes. The phase 3) AHP evaluation produced the application rankings and criterion weights. Phase 3) also gave the method preference, as, immediately after they had used both methods, subjects were asked: "Which method do you prefer to use if you were selecting new software for an international company?"

Outcome attitudes were measured in phase 4), after subjects saw their evaluation outcome. Outcome satisfaction was a multi-item scale validated for generic multicriterion decision processes [77]. It was the response average of the following.

1) I am satisfied with the evaluation outcome (very satisfied–very unsatisfied).
2) How would you rate your satisfaction with the use of the evaluation outcome (very satisfied–very unsatisfied)?
3) Are you satisfied with the use of the evaluation outcome (very satisfied–very unsatisfied)?
4) Considering all things about the evaluation outcome, I am (very pleased–very unpleased).

Further outcome attitude questions were as follows.

1) Confidence: How confident are you that the method's data will give the best software choice? (very confident–very unsure)

TABLE II
SYSTEM ELEMENTS, DECISION MATRIX, AND WEIGHTS

A					Vector weights
	Bndry	Intern Struct	Effect	Recept	
Boundary	1.00	1.02	0.74	0.70	0.21
Internal Structure	0.98	1.00	0.95	0.56	0.21
Effector	1.35	1.35	1.00	0.72	0.26
Receptor	1.42	1.35	1.40	1.00	0.32

CR = 0.005

2) Accurate evaluation: How accurate is the method in testing software performance? (very accurate–very inaccurate)
3) Complete evaluation: The method includes all the dimensions I need in order to evaluate software. (highly agree–highly disagree)

VI. RESULTS

A. Application Rankings

The TAM and WOSP evaluations produced browser ranks of first, second, or third. Half of the subjects ranked one or more browsers differently using TAM versus WOSP criteria, and 27% of the browser ranks changed. To see if the frameworks performed differently, we took the TAM browser rank frequencies as the "expected" and the WOSP browser rank frequencies as "observed," which gave a significant chi-square $(p = 0.012^*)$. We also calculated a rank total for each browser, so if TAM rated the browsers first, second, and third and WOSP rated them third, second, and first, this gave different rank totals. We set the browser with the most changes as the lowest value to minimize the chance of spurious effects. A two-tailed t-test comparison of the WOSP versus TAM rank totals was significant $(p = 0.013^*)$, suggesting that users ranked the browsers significantly differently using the two methods.

H1) *The WOSP application evaluation rankings will differ from the TAM evaluation rankings.*—***supported***

That most subjects stayed with their initial preferences (79% TAM and 72% WOSP) matched earlier findings of a decision "inertia" of about 80% [78]. No one chose MyIE2 as his/her initial preferred browser, but for Explorer and Firefox, the correlation between initial preference and final choice was only borderline for TAM $(p = 0.056, p = 0.045^*)$ but highly significant for WOSP $(p = 0.0015^*, p = 0.0019^*)$, suggesting that WOSP better represented initial user preferences. The average TAM and WOSP rank should be the same, yet an F-test comparison was significant $(p = 0.049^*)$, due to the different number of ties (TAM created 13 ties but WOSP had only 2), suggesting that WOSP has more evaluation discrimination.

B. Criterion Weights

Table II shows the aggregated system element weights $(\text{CR} = 0.005)$, whereas Table III displays the criterion weights $(\text{CR} = 0.000)$. The most valued element was receptor (0.32),

TABLE III
PERFORMANCE-GOAL DECISION MATRIX AND WEIGHTS

Performance Goals		*A*	Vector weights (local)	Vector weights (global)
Boundary	Extendibility	Security		*0.21*
Extendibility	1.00	0.26	0.21	0.04
Security	3.81	1.00	0.79	0.17
CR = 0				
Internal Structure	Flexibility	Reliability		*0.21*
Flexibility	1.00	0.76	0.43	0.09
Reliability	1.32	1.00	0.57	0.12
CR = 0				
Effector	Usability	Functionality		*0.26*
Usability	1.00	1.37	0.58	0.15
Functionality	0.73	1.00	0.42	0.11
CR = 0				
Receptor	Privacy	Connectivity		*0.32*
Privacy	1.00	1.81	0.64	0.21
Connectivity	0.55	1.00	0.36	0.11
CR = 0				

1 followed by effector (0.26), boundary (0.21), and internal structure (0.21).

The most valued criterion was privacy (0.21), followed by security (0.17), usability (0.15), reliability (0.12), connectivity (0.11), and functionality (0.11). Flexibility and extendibility had the lowest ratings, but the weights were spread across all the criteria. For Internet browsers, other criteria are more important than functionality and usability.

H2) *In the WOSP evaluation, other criteria will contribute equally or more than functionality and usability.*—***supported***

1) Criterion Correlations: Most of the WOSP criterion correlations (Table IV) were low, and none was significant. The highest correlation was −0.44 (functionality by security). This suggests that the WOSP model has relatively modular dimensions. Some constructs assumed connected in the research literature were easily distinguished by our users; examples are as follows:

1) privacy (keeping secrets) versus security (protection from attack): correlation of +0.01;
2) extendibility (plug-in/import compatibility) versus connectivity (download ability): correlation of −0.29;
3) usability (reducing effort) versus reliability (maintaining operations): correlation of −0.16.

C. Method Preference

Of 20 subjects, 16 preferred to use the WOSP criteria, 2 preferred TAM, and 2 had no preference. Using AHP, subject judgments were aggregated by geometric means, as recommended [79], to express not only which method is preferred but also by how much it was preferred. Subjects strongly preferred the WOSP model over the TAM model by a factor of 2.56 times, directly after using both methods (Table V).

H3) *Users prefer to use the WOSP rather than the TAM criteria to evaluate software for an organization.*—***supported***

D. Outcome Attitudes

After seeing the evaluation outcome of both methods, subjects were asked what they thought of the methods. Table VI shows that users were more satisfied with the WOSP outcome, were more confident using it, and found it more accurate and complete.

H4) *Users evaluating software for an organization will feel the following:*
- H4a) *more satisfied with the WOSP evaluation outcome*—***supported;***
- H4b) *more confident in their choice using WOSP*—***supported;***
- H4c) *that the WOSP evaluation was more accurate*—***supported;***
- H4d) *that the WOSP evaluation was more complete*—***supported***.

E. Subject Comments

It is not implied that the WOSP criteria are uniformly better than the TAM criteria, as some subjects found TAM simpler and easier to use:

> "WOSP produces better results because it is more precise. TAM is easier and quicker, but the results are less usable. If I were preparing a survey with the ease of taking it in mind only, I would use TAM. Since this is rarely the case, WOSP should be used to generate a more accurate result."

However, others valued the greater detail:

> "I believe that the WOSP method is much better because it forces the subject to look at every single aspect of the system being evaluated, not just the ones that readily come to mind. This ensures that all variables are taken into account when a subject evaluates the system."

F. Conclusions

Software evaluations using the WOSP criteria gave different rankings and fewer ties than the TAM criteria. The WOSP evaluation rated security and privacy (which TAM does not explicitly specify) above functionality and usability. Immediately after using both methods, subjects preferred the WOSP method and were more satisfied with its outcomes. Users could have preferred TAM for its simplicity, or rated functionality and usability the highest, but they did not.

VII. DISCUSSION

A. Possible Limitations

Possible limitations of this paper include the application, the subject number, AHP, and WOSP.

1) Application Context: While the WOSP model applies to any STS, the weights found in this experiment apply specifically to Internet browsers. TAM criterion weights have also been found to vary with application context, e.g., ease of use predicts little in healthcare applications, presumably because healthcare specialists are accustomed to difficult technology [80]. WOSP criterion weights should vary with the situation,

TABLE IV
CRITERION CORRELATIONS

	Ext.	*Sec.*	*Flex.*	*Rel.*	*Usab.*	*Func.*	*Priv.*	*Conn.*
Extendibility	1.00	0.36	-0.20	-0.19	-0.26	-0.32	0.26	-0.29
Security	0.36	1.00	-0.32	-0.26	-0.36	-0.44	0.10	-0.35
Flexibility	-0.20	-0.32	1.00	0.10	-0.07	0.02	-0.41	-0.13
Reliability	-0.19	-0.26	0.10	1.00	-0.16	0.42	-0.31	-0.26
Usability	-0.26	-0.36	-0.07	-0.16	1.00	-0.10	-0.12	0.01
Functionality	-0.32	-0.44	0.02	0.42	-0.10	1.00	-0.35	0.12
Privacy	0.26	0.01	-0.41	-0.31	-0.12	-0.35	1.00	-0.24
Connectivity	-0.29	-0.35	-0.13	-0.26	0.01	0.12	-0.24	1.00

TABLE V
WOSP AND TAM DECISION MATRIX

	A		*Vector weights*
	TAM	*WOSP*	
TAM	1.00	0.39	0.28
WOSP	2.56	1.00	0.72

CR = 0

e.g., a fast-food chain, whose staff changes monthly, may need more usable software than an architecture firm, whose well-trained professionals may prefer hard-to-use but powerful CAD software.

While WOSP criterion weights are expected to vary with application context, they should remain constant for a given context (within experimental error). It is gratifying that a previous browser evaluation study, using different subjects and a different methodology (conjoint analysis), found a similar criterion order, with the same two top and bottom criteria [81].

2) Subject Number: AHP subjects present higher data quality than many other methods, e.g., compare 100 subjects "ticking the boxes" of a questionnaire in, for example, 30 min, with 20 people completing an AHP process at 4 h/person. The 20 AHP subjects represent more data creation effort (80 h) than 100 questionnaire subjects (50 h). Because AHP evaluation requires so much participant time and effort, its supporters argue that it can afford fewer subjects. Also, the significance calculations presented take sample size into account.

3) AHP: AHP requires independent decision criteria that add to a common goal. Both TAM and WOSP have such a common goal—system performance. If the AHP criteria are incomplete, adding another alternative may change the original rankings. AHP also forces decision makers to compare criteria and alternatives, although they often prefer to directly estimate alternatives against criteria [83], whose estimates are quicker but cannot be consistently checked, reducing quality. Despite such limitations, AHP seems a useful tool in multicriterion comparisons.

4) WOSP: Creation cost is a known limit of the WOSP model, which currently has no time dimension—it only represents system performance at a single moment in time. Clearly, cost is a key factor in the decision to purchase or upgrade software, but it may be better to evaluate it separately from system performance, e.g., predefined cost brackets may define quite different types of performance choices. Managerial IS acquisition methods consider tangible and intangible costs and benefits beyond simple purchase cost, like return on investment, cost–benefit total, return on management, and information economics. Multiobjective and multicriterion methods, value analysis methods, and critical success factor methods have advantages and limitations [84]. Combining such established methods with the WOSP model could help managers make better IS/IT evaluations.

B. Implications

1) Risk Versus Opportunity: In this paper, the top four browser selection criteria were the WOSP risk reduction goals of security, privacy, usability, and reliability, suggesting that our subjects saw the Internet today as more of a jungle of danger than a garden of opportunity. This may reflect the rise of online negative forces like viruses, spoofing, spam, phishing, worms, identity theft, pop-up ads, piracy, pornography, and spyware.

2) Extend TAM With WOSP: Subjects evaluating sociotechnical software selected products differently with the two frameworks and preferred the WOSP to the TAM criteria. This suggests that WOSP could augment TAM-based theories, like UTAUT, by replacing the TAM criteria with the WOSP criteria. To replace two criteria (TAM) by eight (WOSP) is a significant increase in theory complexity, but more complex systems may need more complex theories. The WOSP model itself suggests that, as systems develop, more criteria come into play. New systems begin "slack" (with low design tension), but as performance (the web area) increases, so does the design tension. Then, improving one performance dimension can cause another to "bite back" [82]. This could also explain why IS "killer" applications (like e-mail, browsers, or chat) are usually functionally simple (they need the design slack to develop into other performance dimensions).

Conversely, TAM's moderating and contextual variables may carry forward to WOSP, e.g., that gender moderates PU and PEOU (men focus more on usefulness, whereas women concentrate more on ease of use [19]). If men are, in general, more opportunity driven, they may favor the WOSP opportunity criteria (functionality, connectivity, flexibility, and extendibility), whereas if women are more risk averse, they may favor the WOSP criteria (usability, security, reliability, and privacy).

3) Privacy—A Sleeper Social Requirement: The importance of privacy in our subjects' ratings, most of whom were young, suggests that this is a "sleeper" social requirement. The days when Sun Microsystems CEO could say: "You have zero privacy anyway. Get over it." [85] seem over. The quote illustrates the view of technical determinism—that online technical reality

TABLE VI
TAM VERSUS WOSP ATTITUDES

Effect	N	WOSP	TAM	SS	df	MS	F	p
Outcome satisfaction	18	3.14	2.35	5.64	1	5.64	9.6	0.006*
Confidence	20	3.70	2.40	16.9	1	16.9	12.8	0.002*
Accuracy	20	3.60	2.45	13.2	1	13.2	12.4	0.002*
Completeness	19	4.21	2.42	30.4	1	30.4	16.8	0.001*

defines online social activity. The alternate view is that, while technology mediates online society, it does not create it, i.e., human social systems arise from people interacting, and this defines their nature, not the medium. One could consider "real-world" society as a "sociophysical" system, just as one defines an STS. Whether the social interaction occurs via a physical or electronic medium seems a secondary property. Certainly, physics (or technology) can limit social systems, just as, in a 3-ft room, everyone is 3 ft tall, but this is, by constraint, not by nature. If technology disallows privacy, then there is no online privacy. Yet, people may still want choice over their personal information, and social systems that give that individual freedom may still perform better. Personal and social system levels have distinct benefits and requirements, which are not technology determined. Privacy is built into the nature of social connectivity, and it is used by animals as camouflage and by the military as "stealth" systems. People will only find privacy unimportant when it is unimportant how others see us, which will not be any time soon. As the Internet becomes more social, social requirements like privacy will become more important not less.

4) Systems View of Performance and Evaluation: In general, that a lower level (like technology) determines a higher one (like online society) need not make it the best way to run the system, e.g., line voltages (hardware) determine software code entirely, but writing code using line voltages is foolish. The higher software level offers advantages over the hardware level, and if its new requirements are met, it can increase system performance (assuming that hardware requirements are also met). Similarly, social synergy offers enormous productivity increases, given that social requirements like privacy and fairness are met. The conclusion is that online prosperity will increase as human and social requirements drive technical design [86].

If one graphs the functionality and usability criteria, their "efficient frontier" is a line that defines the best that one can achieve of one criterion for a given value of the other [87]. To design a system that is both functional and usable requires a "synthesis of form," which reconciles the different requirements [88]. The efficient frontier concept implies that there are many "best combinations" of functionality and usability, not one. If performance were a single dimension, a single point could represent the best design. In a 2-D space, a line represents the efficient frontier. If, as the WOSP model suggests, there are at least eight significant performance dimensions, the efficient frontier is itself a multidimensional space, perhaps why software design is as much an art as a science.

The WOSP model's implications can be summarized as follows.

1) System performance is multidimensional—no single design criterion describes all performance requirements, e.g., functionality, usability, reliability, flexibility, security, extendibility, connectivity, and privacy can all be critical.
2) System performance depends on the environment—design criterion weights (0%–100%) change with the environment, e.g., security has more weight in a threat environment.
3) Improving one performance criterion can reduce another, as all criteria involve the same system architecture, e.g., making a network more secure can make it less usable.
4) The efficient frontier of "best performance" has many dimensions—there are many "best" designs, e.g., slow armored vehicles and fast unarmored ones both perform well in different ways.
5) Innovations can expand the efficient frontier by reconciling criterion conflicts, e.g., lightweight armor that allows fast and armored vehicles.
6) Different system levels invoke different criteria—each level redefines what "the system" is, e.g., hardware reliability, software reliability, user reliability, and community reliability are different.
7) Higher system levels offer higher performance at the cost of higher added requirements, e.g., social synergy improves e-trade if the social requirement of fairness enables social stability.

The WOSP model expands current concepts of system performance in two ways. First, it gives NFR and quality criteria a home within the general concept of "performance." It avoids statements like: "Flexibility or performance? That choice is a constant trade-off for micro-processor designers" [90, p. 58]. Why should one trade off performance against other things? In the WOSP view, flexibility can trade off with reliability or functionality, etc., but not with performance because flexibility is part of performance. Second, the WOSP performance concept can work at different IS levels, and so, it can explain how technically perfect systems can still fail on social levels. This vision of performance as having many dimensions and many levels is more complex than previous models. Yet, the enormous variety of "success," both in biological life and modern software [89], suggests that this view is not unreasonable. The WOSP advantages are the following.

1) It offers a useful checklist of potentially critical criteria for new software designers and adopters. Given that systems fail in more ways than they succeed, this seems important.

2) It accommodates how criteria change as systems evolve higher IT/IS levels. Given that modern software is rapidly becoming more social, this also seems important.
3) Finally, like the simpler TAM, it has system design roots, so all valid TAM research can carry forward. Given the cost of research, this is also important.

The WOSP limits are the following.

1) It currently lacks a time dimension, and so, it must be used with a cost estimate method.
2) Its criteria are general and must be specifically defined for each application (see Appendix I).
3) Criterion weights vary with application context.
4) The IT/IS level of use must be predefined.

However, it is our wish that, overall, the WOSP extension to TAM will help designers, users, and managers better design, operate, and evaluate modern sociotechnical software.

Appendix I
TAM/WOSP Criteria and Statements

A. TAM

The TAM criteria and grounding statements were as follows [18].

1) Usefulness (PU): That the software is useful.
   a) Using the browser increases my productivity.
   b) Using the browser increases my job performance.
   c) Using the browser enhances my effectiveness on the job.
2) Ease of use (PEOU): That the software is easy to use.
   a) Learning to operate this browser is easy for me.
   b) I find it easy to get the this browser to do what I want to do.
   c) This browser is not rigid and inflexible to interact with.

B. WOSP

The WOSP selection criteria were first broken down by the following four system elements:

1) boundary (third-party use/abuse): to permit, or deny, other third-party programs or data to enter the system;
2) internal structure (changes/contingencies): to manage unexpected changes or events inside or outside the system;
3) effector (output efficiency): to produce some task output in an efficient way;
4) receptor (communications): to connect to other people or systems.

These gave eight subcriteria statements and three grounding statements each, based on the WOSP model, as follows.

1) Extendibility: The ability to make use of third-party programs and data.
   a) It works with all third-party multimedia tools, like Real media player and Flash.
   b) It follows all World Wide Web source code and data standards, e.g., unicode.
   c) It can handle graphics, sound, and video in a wide variety of different formats.
2) Security: The ability to defend against hostile attack, unauthorized entry, damage, hurt, or takeover.
   a) When a file is downloaded to the hard drive, it is checked for viruses before use.
   b) It can detect and prevent spyware from installing.
   c) It can detect and prevent pop-up ads.
3) Reliability: The ability to continue working despite errors/problems or to quickly recover from failure.
   1) It never breaks down or "hangs" (fails to respond), even if I use it for a long time.
   2) If one part of the browser fails, like a plug-in, the entire browser does not crash.
   3) Even if I multitask, and do many things at once, it still works well.
4) Flexibility: The ability to change itself to fit different situations.
   a) It runs on all our computers and operating systems.
   b) It is easily changed to fit disability needs, e.g., larger text or graphics for those with poor vision.
   c) It has a preference "control panel" to change browser settings.
5) Functionality: The ability to move to Web sites and display their information.
   a) The Favourites list lets me jump directly to my favorite sites.
   b) This browser gets me where I want to go quickly.
   c) The browser has everything I need to search, navigate, and display the Internet.
6) Usability: The ability to be used easily.
   a) The user interface is consistent and easy to learn.
   b) I did not need training to use it the first time.
   c) I accomplish my tasks easier and quicker with this browser.
7) Connectivity: The ability to exchange information with other systems.
   a) When downloading, it gives useful information, like the estimated time to complete the download.
   b) If a download is stopped, or fails for any reason, it can be restarted again later from where it left off, saving a lot of time.
   c) It gives access to other ways of communicating, like Telnet, Ftp, e-mail, and chat.
8) Privacy: The ability to limit unwanted information disclosure.
   a) Any sensitive information I give the browser, like log-on passwords, is encrypted, so others cannot see it.
   b) Password information always shows as asterisks, so others cannot look over my shoulder to see them.
   c) It stops Web sites from getting my name or e-mail from my computer's data.

References

[1] J. C. Westland and T. H. K. Clark, *Global Electronic Commerce*. Cambridge, MA: MIT Press, 2000.
[2] P. B. Seddon, V. Graeser, and L. P. Willcocks, "Measuring organizational IS effectiveness: An overview and update of senior management perspectives," *ACM SIGMIS Database*, vol. 33, no. 2, pp. 11–28, 2002.
[3] S. Taylor and P. A. Todd, "Assessing IT usage: The role of prior experience," *MIS Q.*, vol. 19, no. 4, pp. 561–570, Dec. 1995.

[4] E. Brynjolfsson, "The productivity paradox of information technology," *Commun. ACM*, vol. 36, no. 12, pp. 66–77, Dec. 1993.
[5] M. Y. Ivory and M. A. Hearst, "The state of the art in automating usability evaluation of user interfaces," *ACM Comput. Surv.*, vol. 33, no. 4, pp. 470–516, Dec. 2001.
[6] W. J. Orlikowski and D. G. Gash, "Technological frames: Making sense of information technology in organizations," *ACM Trans. Inf. Syst.*, vol. 12, no. 2, pp. 174–207, Apr. 1994.
[7] J. R. Ralston, *Voltaire's Bastards*. New York: Penguin Books, 1993.
[8] B. N. Reeves and A. C. Lemke, "The problem as a moving target in cooperative system design," in *Proc. HCI Consortium Workshop*, Jan. 1991.
[9] J. Grudin, "Human factors, CHI and MIS," in *Advances in Manage. Inf. Syst.*, P. Zhang and D. Galletta, Eds. Armonk, NY: M. E. Sharpe, 2005.
[10] M. Fishbein and I. Ajzen, *Belief, Attitude, Intention and Behaviour: An Introduction to Theory and Research*. Reading, MA: Addison-Wesley, 1975.
[11] F. D. Davis and V. Venkatesh, "A critical assessment of potential measurement biases in the technology acceptance model: Three experiments," *Int. J. Human-Comput. Stud.*, vol. 45, no. 1, pp. 19–45, Jul. 1996.
[12] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *Manage. Inf. Syst. Q.*, vol. 13, no. 3, pp. 319–340, Sep. 1989.
[13] D. Hix and R. S. Schulman, "Human–computer interface development tools: A methodology for their evaluation," *Commun. ACM*, vol. 34, no. 3, pp. 75–87, Mar. 1991.
[14] J. Lin and H. Lu, "Towards an understanding of the behavioral intention to use a web site," *Int. J. Inf. Manage.*, vol. 20, no. 3, pp. 197–208, Jun. 2000.
[15] L. Vijayasarathy, "Predicting consumer intentions to use on-line shopping: The case for an augmented technology acceptance model," *Inf. Manage.*, vol. 41, no. 6, pp. 747–762, Jul. 2004.
[16] V. Lai and H. Li, "Technology acceptance model for internet banking: An invariance analysis," *Inf. Manage.*, vol. 42, no. 2, pp. 373–386, Jan. 2005.
[17] H. Heijden, "Factors influencing the usage of websites: The case of a generic portal in The Netherlands," *Inf. Manage.*, vol. 40, no. 6, pp. 541–549, Jul. 2003.
[18] P. Legris, J. Ingham, and P. Collerette, "Why do people use information technology? A critical review of the technology acceptance model," *Inf. Manage.*, vol. 40, no. 3, pp. 191–204, Jan. 2003.
[19] V. Venkatesh and M. G. Morris, "Why don't men ever stop to ask for directions? Gender, social influence, and their role in technology acceptance and usage behavior," *MIS Q.*, vol. 24, no. 1, pp. 115–139, Mar. 2000.
[20] I. Brown, "Individual and technological factors affecting perceived ease of use of web-based learning technologies in a developing country," *Electron. J. Inf. Syst. Dev. Ctries.*, vol. 9, no. 5, pp. 1–15, 2002.
[21] V. Venkatesh, "Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model," *Inf. Syst. Res.*, vol. 11, no. 4, pp. 342–365, Dec. 2000.
[22] J. Moon and Y. Kim, "Extending the TAM for a World-Wide-Web context," *Inf. Manage.*, vol. 38, no. 4, pp. 217–230, Feb. 2001.
[23] C. Ong, J. Lai, and Y. Wang, "Factors affecting engineer's acceptance of asynchronous e-learning systems in high-tech companies," *Inf. Manage.*, vol. 41, no. 6, pp. 795–804, Jul. 2004.
[24] H. Shih, "An empirical study on predicting user acceptance of e-shopping on the Web," *Inf. Manage.*, vol. 41, no. 3, pp. 351–368, Jan. 2004.
[25] J. Yu *et al.*, "Extending the TAM for a t-commerce," *Inf. Manage.*, vol. 42, no. 7, pp. 965–976, Oct. 2005.
[26] I. Benbasat and H. Barki, "Quo Vadis TAM," *J. Assoc. Inf. Syst.*, vol. 8, no. 4, pp. 212–218, Apr. 2007.
[27] L. Festinger, *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford Univ. Press, 1957.
[28] V. Venkatesh and M. G. Morris, "User acceptance of information technology: Toward a unified view," *MIS Q.*, vol. 27, no. 3, pp. 425–478, Sep. 2003.
[29] S. E. Asch, *Social Psychology*. New York: Prentice-Hall, 1952.
[30] R. S. Crutchfield, "Conformity and character," *Amer. Psychol.*, vol. 10, pp. 191–198, 1955.
[31] M. Lea and R. Spears, "Computer-mediated communication, de-individuation and group decision-making," *Int. J. Man-Mach. Stud.*, vol. 34, no. 2, pp. 283–301, Feb. 1991.
[32] T. Postmes, "Social influence in computer-mediated groups," Ph.D. dissertation, Univ. Amsterdam, Amsterdam, The Netherlands, 1997.
[33] B. Whitworth, B. Gallupe, and R. McQueen, "Generating agreement in computer-mediated groups," *Small Group Res.*, vol. 32, no. 5, pp. 625–665, 2001.
[34] B. Whitworth, "Polite computing," *Behav. Inf. Technol.*, vol. 24, no. 5, pp. 353–363, Sep. 2005. [Online]. Available: http://brianwhitworth.com/polite.rtf
[35] P. J. Hu *et al.*, "Examining the technology acceptance model using physician acceptance of telemedicine technology," *J. Manage. Inf. Syst.*, vol. 16, no. 2, pp. 91–112, Sep. 1999.
[36] I. Sommerville, *Software Engineering*, 7th ed. Upper Saddle River, NJ: Pearson Educ., 2004.
[37] D. Compeau, C. A. Higgins, and S. Huff, "Social cognitive theory and individual reactions to computing technology: A longitudinal study," *MIS Q.*, vol. 23, no. 2, pp. 145–158, Jun. 1999.
[38] G. Moore and I. Benbasat, "Development of an instrument to measure the perceptions of adopting an information technology innovation," *Inf. Syst. Res.*, vol. 2, no. 3, pp. 192–222, Sep. 1991.
[39] K. E. Boulding, "General systems theory—The skeleton of science," *Manage. Sci.*, vol. 2, no. 3, pp. 197–208, Apr. 1956.
[40] J. Grudin, "The computer reaches out: The historical continuity of user interface design," in *Proc. CHI, ACM SIGCHI Conf.*, Seattle, WA, 1990, pp. 261–268.
[41] K. Kuutti, "Activity theory as a potential framework for human–computer interaction," in *Research Context and Consciousness: Activity Theory and Human–Computer Interaction*, B. A. Nardi, Ed. Cambridge, MA: MIT Press, 1996.
[42] S. Alter, "A general, yet useful theory of information systems," *Commun. AIS*, vol. 1, no. 3, pp. 13–60, Mar. 1999.
[43] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Urbana, IL: Univ. Illinois Press, 1949.
[44] B. Whitworth, "Social-technical systems," in *Encyclopedia of Human Computer Interaction*, C. Ghaoui, Ed. London, U.K.: Idea Group Reference, 2006, pp. 533–541.
[45] B. Whitworth, B. Gallupe, and R. J. McQueen, "A cognitive three-process model of computer-mediated group interaction," *Group Decis. Negot.*, vol. 9, no. 5, pp. 431–456, Sep. 2000.
[46] F. Losavio *et al.*, "Designing quality architecture: Incorporating ISO standards into the unified process," *Inf. Syst. Manage.*, vol. 21, no. 1, pp. 27–44, Dec. 2004.
[47] T. Berners-Lee, *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web*. New York: Harper-Collins, 2000.
[48] A. Gargaro *et al.*, "The power of standards," *Commun. ACM*, vol. 36, no. 8, pp. 11–12, Aug. 1993.
[49] M. De Simone and R. Kazman, "Software architectural analysis: An experience report," in *Proc. CASCON*, Toronto, ON, Canada, 1995.
[50] K. Knoll and S. L. Jarvenpaa, "Information technology alignment or 'fit' in highly turbulent environments: The concept of flexibility," in *Proc. SIGCPR*, Alexandria, VA, 1994.
[51] P. Regan, *Legislating Privacy, Technology, Social Values and Public Policy*. Chapel Hill, NC: Univ. North Carolina Press, 1995.
[52] J. C. Laprie and A. Costes, "Dependability: A unifying concept for reliable computing," in *Proc 12th Int. Symp. Fault-Tolerant Comput.*, 1982, pp. 18–21.
[53] *Guidelines for the Security of Information Systems*, OECD, Paris, France, 1996, p. 50.
[54] E. Jonsson, "An integrated framework for security and dependability," in *Proc. NSPW*, Charlottsville, VA, 1998, pp. 22–29.
[55] G. Gediga, K. Hamborg, and I. Duntsch, "The IsoMetrics usability inventory: An operationalization of ISO9241-10 supporting summative and formative evaluation of software systems," *Behav. Inf. Technol.*, vol. 18, no. 3, pp. 151–164, 1999.
[56] B. Nuseibeh and S. Easterbrook, *Requirements Engineering: A Roadmap, in the Future of Software Engineering*, A. Finkelstein, Ed. Limerick, Ireland: ACM Press, 2000.
[57] J. G. Walls, G. R. Widmeyer, and O. A. El Sawy, "Building and information system design theory for vigilant EIS," *Inf. Syst. Res.*, vol. 3, no. 1, pp. 36–59, Mar. 1992.
[58] A. R. Hevner, S. T. March, and J. Park, "Design science in information systems research," *MIS Q.*, vol. 28, no. 1, pp. 75–105, 2004.
[59] L. V. Bertalanffy, *General System Theory*. New York: George Braziller Inc., 1968.
[60] L. Chung *et al.*, *Non-functional Requirements in Software Engineering*. Boston, MA: Kluwer, 1999.
[61] B. Whitworth, J. Fjermestad, and E. Mahinda, "The web of system performance: A multi-goal model of information system performance," *Commun. ACM*, vol. 49, no. 5, pp. 93–99, May 2006.
[62] B. Whitworth and M. Zaic, "The WOSP model: Balanced information system design and evaluation," *Commun. Assoc. Inf. Syst.*, vol. 12, pp. 258–282, 2003.

[63] K. Patten *et al.*, “Leading IT flexibility: Anticipation, agility and adaptability,” in *Proc. Amer. Conf. Inf. Syst.*, Omaha, NE, Aug. 11–14, 2005.

[64] O. A. El Sawy and B. Nanus, “Towards the design of robust information systems,” *J. Manage. Inf. Syst.*, vol. 5, no. 4, pp. 33–54, 1989.

[65] B. A. Nixon, “Managing performance requirements for information systems,” in *Proc. 1st Int. Workshop Softw. Perform.*, Santa Fe, NM, 1998, pp. 131–144.

[66] N. S. Rosa, G. R. R. Justo, and P. R. F. Cunha, “A framework for building non-functional software architectures,” in *Proc. SAC*, Las Vegas, NV, 2001, pp. 141–147.

[67] L. M. Cysneiros and J. Leita, “Non-functional requirements: From elicitation to modelling languages,” in *Proc. ICSE*, Orlando, FL, 2002, pp. 699–700.

[68] J. K. Pinto, “Project management 2002,” *Res. Technol. Manage.*, vol. 45, no. 2, pp. 22–37, Mar./Apr. 2002.

[69] D. M. Kienzle and W. A. Wulf, “A practical approach to security assessment,” in *Proc. New Security Paradigms Workshop*, Cumbria, U.K., 1997, pp. 5–16.

[70] N. S. Borenstein, “Power, ease of use and cooperative work in a practical multimedia message system,” *Int. J. Man-Mach. Stud.*, vol. 34, no. 2, pp. 229–259, Feb. 1991.

[71] B. Whitworth and E. Whitworth, “Reducing spam by closing the social–technical gap,” *Computer*, vol. 37, no. 10, pp. 38–45, Oct. 2004.

[72] A. Moreira, J. Araujo, and I. Brita, “Crosscutting quality attributes for requirements engineering,” in *Proc. SEKE*, Ischia, Italy, 2002, pp. 167–174.

[73] W. H. De Lone and E. R. McLean, “Information systems success: The quest for the dependent variable,” *Inf. Syst. Res.*, vol. 3, no. 1, pp. 60–95, Mar. 1992.

[74] I. Benbasat and R. W. Zmud, “The identity crisis within the IS discipline: Defining and communicating the discipline’s core properties,” *MIS Q.*, vol. 27, no. 2, pp. 183–194, Jun. 2003.

[75] T. L. Saaty, “A scaling method for priorities in hierarchical structures,” *J. Math. Psychol.*, vol. 15, no. 3, pp. 234–281, 1977.

[76] T. L. Saaty, *The Analytic Hierarchy Process*. New York: McGraw-Hill, 1980.

[77] Q. Zhang, J. Chen, and P. Chong, “Decision consolidation: Criteria weight determination using multiple preference formats,” *Decis. Support Syst.*, vol. 38, no. 2, pp. 247–258, Nov. 2004.

[78] B. Whitworth and R. Felton, “Measuring disagreement in groups facing limited choice problems,” *Database Adv. Inf. Syst.*, vol. 30, no. 3/4, pp. 22–33, 1999.

[79] T. L. Saaty and L. G. Vargas, *Models, Methods, Concepts & Applications of the Analytic Hierarchy Process*. Boston, MA: Kluwer, 2001.

[80] W. G. Chismar and S. Wiley-Patton, “Does the extended technology acceptance model apply to physicians?” in *Proc. 36th Hawaii Int. Conf. Syst. Sci.*, Big Island, HI, 2003, p. 160.

[81] E. Mahinda and B. Whitworth, “The web of system performance: Extending the TAM model,” in *Proc. AMCIS*, Omaha, NE, Aug. 11–14, 2005.

[82] E. Tenner, *Why Things Bite Back*. New York: Vintage Books, 1997.

[83] C. Sylla and B. Arinze, “A method for quality precoordination in a quality assurance information system,” *IEEE Trans. Eng. Manage.*, vol. 38, no. 3, pp. 245–256, Aug. 1991.

[84] C. Sylla and J. Wen, “A conceptual framework for evaluation of information technology investments,” *Int. J. Technol. Manage.*, vol. 24, no. 2/3, pp. 236–261, 2002.

[85] S. Manes, “Private lives? Not ours!” *PC World*, vol. 18, no. 6, p. 312, Jun. 2000.

[86] B. Whitworth and A. deMoor, “Legitimate by design: Towards trusted virtual community environments,” *Behav. Inf. Technol.*, vol. 22, no. 1, pp. 31–51, 2003.

[87] R. L. Keeney and H. Raiffa, *Decisions With Multiple Objectives: Preferences and Value Tradeoffs*. New York: Wiley, 1976.

[88] C. Alexander, *Notes on the Synthesis of Form*. Cambridge, MA: Harvard Univ. Press, 1964.

[89] J. S. David, W. E. McCarthy, and B. S. Sommer, “Agility—The key to survival of the fittest,” *Commun. ACM*, vol. 46, no. 5, pp. 65–69, May 2003.

[90] R. Koch, *Flexibility or Performance?* Harlan, IA: Scientific Amer., Aug. 2005, p. 58.

Brian Whitworth received the B.Sc. degree in mathematics, the B.A. degree in psychology, the M.A. degree (first class) in neuropsychology, and the Ph.D. degree in information systems from Waikato University, Hamilton, New Zealand.

He is currently a Senior Lecturer with the Institute of Information and Mathematical Sciences, Massey University at Albany, Auckland, New Zealand. He has contributed to journals like *Small Group Research*, *Group Decision and Negotiation*, the *Database Journal*, *Communications of the AIS*, IEEE COMPUTER, *Behavior and Information Technology (BIT)*, and *Communications of the ACM*. His main interest is the design, operation, and evaluation of sociotechnical systems, including email, bulletin boards, chat, online markets, social networks, and social knowledge exchange systems. He is the main editor of the *Handbook of Research on Socio-Technical Design and Social Networking Systems* due in 2009. He was Lecturer of the Year for 2006 for the Massey University Albany Campus.

Victor Bañuls was born in Huelva, Spain, in 1977. He received the B.S. degree in business administration from the University of Huelva, Huelva, Spain, and the M.S. degree in computer systems.

He is currently an Assistant Professor of management information systems with the Department of Management, Pablo de Olavide University, Seville, Spain. Before his current position, he was an Information Systems Consultant and has participated in several international research projects. He also served as a Visiting Research Scholar at the Department of Information Systems, New Jersey Institute of Technology, Newark. His research has been published in journals such as *Technological Forecasting and Social Change*, *Futures*, and *Technovation*. His current main research interests are information systems assessment and technology foresight methodologies.

Cheickna Sylla received the Ph.D. degree in industrial engineering from the State University of New York, Buffalo.

He is currently an Associate Professor of management science and decision science with the School of Management, New Jersey Institute of Technology (NJIT), Newark. He is the author of over 60 articles published in the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, the IEEE TRANSACTIONS ON ENGINEERING MANAGEMENT, the *European Journal of Operations Research*, *Human Factors*, *Decision Sciences*, *Computers and Industrial Engineering*, *Control Engineering Practice*, *Cost Management*, the *International Journal of Management and Decision Making (IJMDM)*, the *International Journal of Technology Management (IJITM)*, the *International Journal of Information and Technology Management*, and numerous conference proceedings. He was a Guest Editor of the *International Journal of Management and Decision Making* and the *International Journal of Information Technology and Management*. His teaching and research interests include decision support and information systems modeling and evaluation, statistical analyses and applied operations research methods to human-machine interface studies, project management, supply chain and distribution logistics, and operations management.

Dr. Sylla is a Senior Member of the Institute for Operations Research and the Management Sciences, the Institute of Industrial Engineers, and the Production and Operations Management Society. He is the recipient of the Best Paper Awards at the 1992 International Conference on Information Systems and the 1995 Academy of Management Conference. He also received the Van Houten Distinguished Teaching Award in 2003 and the Cullimore Distinguished Service Award in 2001 for his professional contributions at NJIT.

Edward Mahinda, photograph and biography not available at the time of publication.