

Extending 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 socio-technical software. Research on the Technology Acceptance Model (TAM) confirms that perceived usefulness (PU) and perceived ease of use (PEOU) 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 study compares WOSP and TAM criterion frameworks in a performance evaluation experiment using the Analytic Hierarchy Process (AHP) 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 socio-technical software becomes more complex, users may need (or prefer) more comprehensive evaluation criteria frameworks.

Index Terms—System performance, technology assessment, software requirements, socio-technical

I. INTRODUCTION

¹Over the last two decades, over 50% of new organizational capital investments have been in IT [1]. The 2001 total worldwide IT expenditure exceeded one trillion US 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 information technology (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 multi-billion 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:

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 to 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. Part II reviews current technology evaluation frameworks, Part III introduces a new criterion framework, Part IV presents a strategy to compare criterion frameworks, Part V gives the method, Part VI analyzes the results, and Part VII discusses limitations and implications. The criteria and statements used are given in Appendix A.

II. CURRENT TECHNOLOGY EVALUATION CRITERIA

A. The Perspectives

Three perspectives seem to permeate the study of people and technology, each with a different research culture, different journals and different conferences, summarized as [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,

¹ Manuscript received Nov 25, 2005.

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

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

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

Edward Mahinda is with the Information Systems Department, New Jersey Institute of Technology, USA

This paper was partly funded by grants from Fulbright Foundation, Washington, D.C., and by NJIT Research Office.

giving the question: “*What do users want from software?*”

2. *Computer Human Interaction (CHI) 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 (MIS) 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 end-user and the IT manager. That introducing new technology involves designers, users and managers, suggests that its evaluation criteria should be meaningful to all, as:

1. *Criteria useful to users only*, like “attractiveness”, let users know it 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, the users who use IT, and the managers who fund IT.*

B. Technology Acceptance Model

The Technology Acceptance Model (TAM) derives from the Theory of Reasoned Action (TRA) [10], and

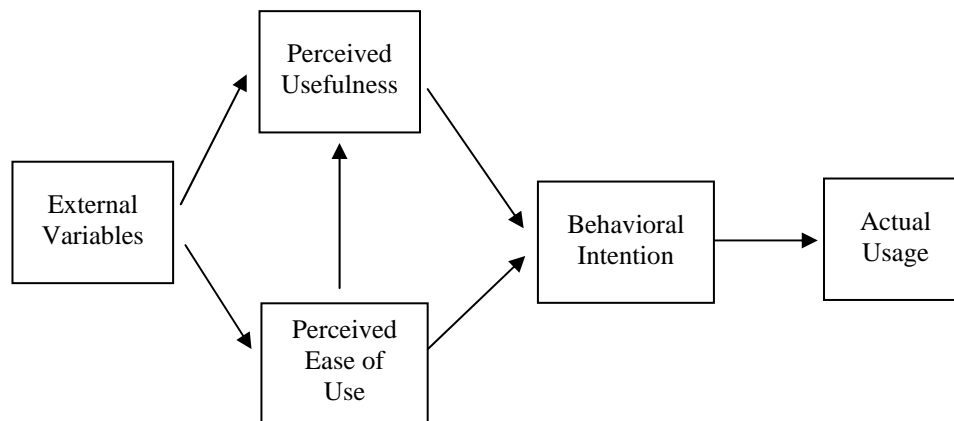


Figure 1. The Technology Acceptance Model

is shown in Figure 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 bi-dimensional, with usability and usefulness 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.*” [18, p202]. Recognition of TAM’s incompleteness has increased over the years since its inception, leading to many suggested extensions.

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, researcher sought to extend TAM using cognitive variables that predicted also acceptance, like *playfulness* [22], *credibility* [23], *attractiveness* [17], *self-efficacy* [23], *behavioural 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, p2]

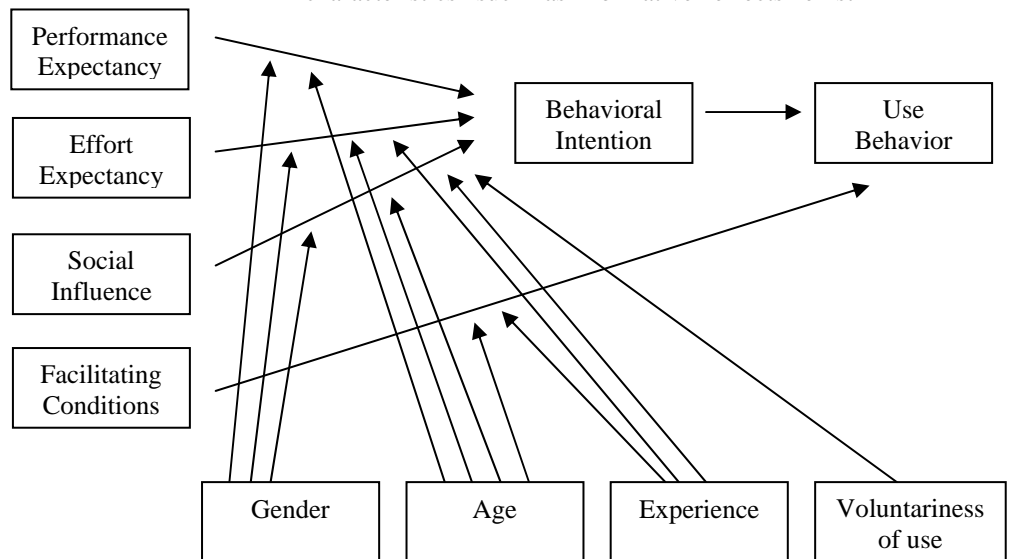
One problem with psychological criteria is that perceptions gathered from the same subjects at the same time can confound because they originate in the same mind, e.g. computer anxiety may reduce ease of use, but hard to use software may also cause anxiety. In cognitive dissonance theory people mutually 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 do 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 two-dimensional 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 (Figure 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 two well known organizational constructs, namely:

1. *Social influence*, the degree users believe important others support the system,
2. *Facilitating technology*, the degree users believe organizational and technical infrastructure exists to support the system.

Social influence suggests users will accept new technology when significant others have already done so. Such “conformity” has a long history in social research [29, 30, 31-33]. Facilitating technology suggests 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 don’t explain how new technology takes hold in the first place.



New products, like blue-tooth, by definition often begin without either infrastructure or normative support. What distinguishes products like cell-phones, which take off, from those that don’t, like the vid-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, 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 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

Figure 2. The UTAUT Model [28]

sociology models, like innovation diffusion theory [38]. Yet these models are all relevant to IS/IT evaluation. One way to connect them conceptually is to postulate IS “levels” to match the variable types above [39]. Grudin has suggested three levels: hardware, software and cognitive [40], while Kuutti adds a social work processes level [41]. We also suggest *four system levels*:

Level 1. Hardware: computers, wires, printer, keyboard, mouse

Level 2. Software: programs, data, bandwidth, memory

Level 3. Human-Computer Interaction (HCI): Attitudes, beliefs, opinions, ideas

Level 4. Socio-technical. Norms, culture, laws, sanctions, roles

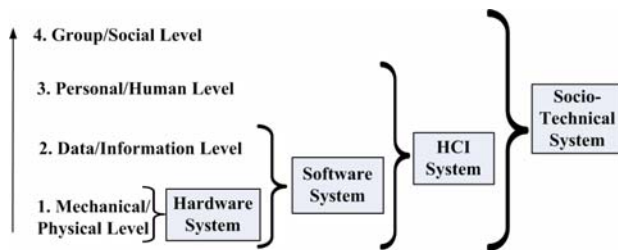


Figure 3. IS/IT Levels and Systems

These four levels match the disciplines of Engineering, Computing, Psychology and Social Sciences, respectively. They also match Alter’s view of IS as hardware, software, people and processes [42]. They combine to create four system types (see Figure 3). Hardware and software gives *technology*. People plus technology gives *human-computer interaction* systems (hardware + software + people), which can exchange meaning as well as data or information (where “information” is as Shannon and Weaver define it [43]). Finally, when individuals form communities mediated by technology, the result is a four level *socio-technical 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 above [33, 49].

In this view, a socio-technical system is a social system built upon a technical system. One could equally consider normal social interaction as a “socio-physical 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 creates 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 socio-technical 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, so: “... *repeatedly demonstrating that certain mediators (beliefs) are influential without understanding how to influence such beliefs through IT design is ultimately of limited value.*” [26, p7].

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 [61]. They argue that if IS research doesn’t contribute to IS design [62] 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, p24]. However the ISO 9126-1 quality model suggests functionality, usability, reliability, efficiency, maintainability and portability as critical criteria [50]. Berners-Lee however found scalability the key to World Wide Web success [51], while others espouse open standards [52]. Alter suggests the criteria of cost, quality, reliability, responsiveness and conformance to standards [42]. Software architects argue for portability, modifiability and extendibility [53], while others find flexibility a critical success factor [54]. Still others suggest privacy is what users really want [55]. On the issue of what

criteria human-machine systems must 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 [57]. This makes reliability an aspect of security. However “Dependability” has been defined as reliability, availability, safety and security [56], making security part of a larger reliability type construct. Is reliability part of security, or vice-versa? Since 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 [58]. 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 [59], another review finds “scalability”, “robustness” and “connectivity” aspects of an equally general “flexibility” concept [54, p6]. 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...*” [60].

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. A GENERAL SYSTEMS APPROACH

The Web of System Performance (WOSP) model uses a general systems perspective [63] to define performance, and decompose it into a multi-goal model as suggested by Chung [64]. A brief description follows, as a detailed justification is given elsewhere [65, 66]. The model assumes only a general “system”, so 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. *The Web of System Performance*

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” respectively exist, then information systems, cognitive systems and social systems 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 four system elements:

1. *Boundary*: Separates the system from the environment.
2. *Internal structure*: Supports and coordinates the system
3. *Effectors*: Acts upon the environment.
4. *Receptors*: Analyses 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 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 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 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 1 are new, as similar concepts pervade the IS/IT literature, but this model integrates them into a common framework (Figure 4), where:

- *The web area* is the overall system performance. A larger area has more performance potential.
- *The web shape* shows each goal’s contribution to performance in a given environment, e.g. turbulent environments favour reliability.

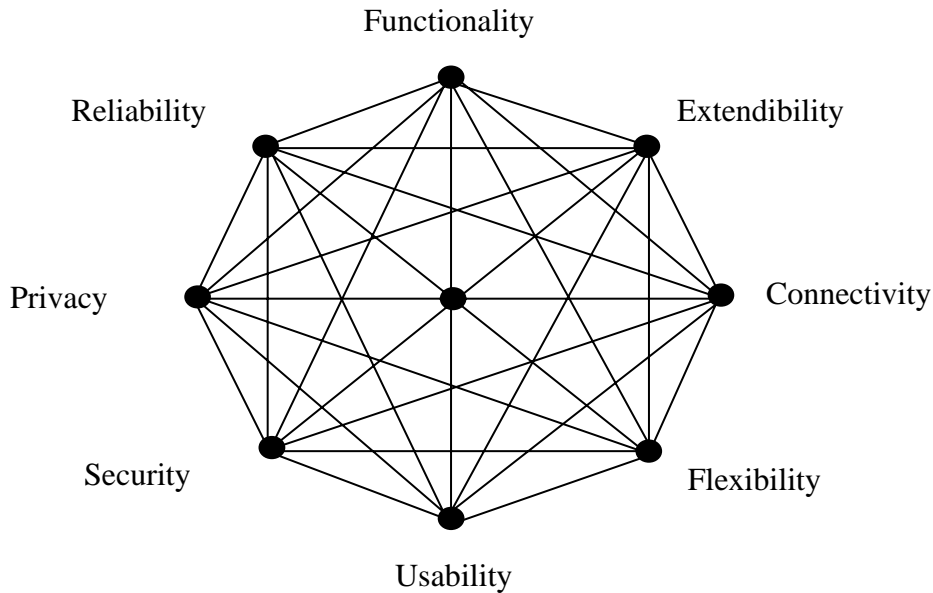


Figure 4. The Web of System Performance

- *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. *Multi-level applicability.* Criteria can apply at any IS level (but only to one level at a time).

Criteria Validity

Unlike models that simply state relevant constructs, the WOSP model defines its general concern (performance), then derives criteria. The results match well with known criteria, e.g. extendibility matches with openness, interoperability, permeability, compatibility, and scalability [51]. Flexibility matches adaptability, agility, portability, customizability, plasticity and modifiability [67]. 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.

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" [69]. For decades these "-ilities" have stood apart from the main "what the system does" specification. They have defied categorization, as while most agree they are important, there is little agreement on how they relate to performance (or each other). The WOSP model queries the assumption that "non-functional" requirements (NFRs) differ essentially from functional ones [70], as it does not distinguish "functional" and "non-functional" performance goals.

Consider the argument that NFR's 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. The same logic that makes reliability secondary to functionality also makes functionality secondary to reliability. Nor is it easy to argue that functionality is "obviously" more important, as many software failures involve NFRs [71, p699], and the error modules that create reliability often involve more code 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

Table 1. System Performance Goals

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

performance as functionality. WOSP has four risk reducing criteria (security, reliability, privacy and usability) and four opportunity enhancing criteria (functionality, extendibility flexibility and connectivity). Since reducing risk is as important to system performance as increasing success [72], 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.

Criteria Modularity

Criteria modularity means that the criterion goals do not overlap or contradict in principle. As already noted, the goals of Table 1 interact in design tensions, but criteria modularity means they do not inherently contradict. If achieving one criteria need not necessarily prevent another, all combinations are possible, e.g. a bullet-proof plexi-glass 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, the first aiming to provide services and the second to deny services [58]. 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. So functionality is not the inevitable enemy

of usability [74], nor does connectivity mean the end of privacy [75].

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 [68]. In the system requirements literature, such interacting goals are “cross-cutting requirements” [76], 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.” [73]

Hence designers must satisfy not only the eight Table 1 criteria, but also their interactions. Fortunately, apparent opposites like flexibility and reliability are goals in tension not simple trade-offs, and can be “reconciled” by innovative design.

Multi-level Applicability

The WOSP model can apply to any IS level, e.g. a system can be hardware reliable but software unreliable, or both hardware and software reliable but operator unreliable [36, p24]. 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 “meta-theory” 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 socio-technical, which suits the WOSP model, with its social dimension (connectivity-privacy). While individuals evaluating technology for

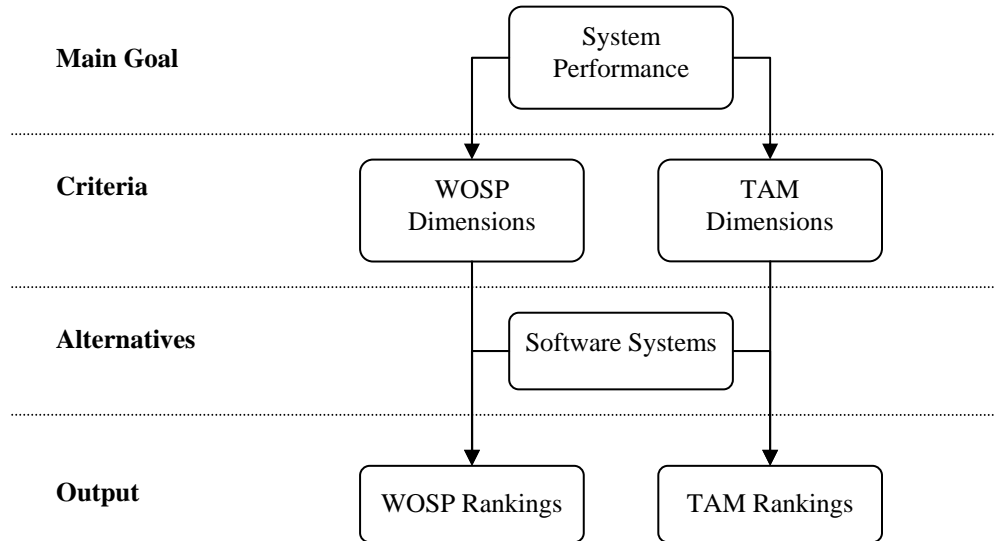


Figure 5. AHP Design

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 [77]. The WOSP model suggests the criteria *perceived* security, extendibility, flexibility, reliability, functionality, usability, connectivity and privacy, while TAM suggests perceived usefulness and usability. The research goal is to compare these two evaluation criteria 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* it's now very general "usefulness" construct. Otherwise research showing that "usefulness is useful" is like finding

better performance is better, which may add little value [78]. However this second case creates an inconsistency - if NFRs like security fall under a generic "usefulness" construct, why doesn't the same logic apply to ease of use? Isn't ease of use 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 study 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 different software evaluations:

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

Since 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.

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:

H4a. More satisfied with the WOSP evaluation outcome.

H4b. More confident in their choice using WOSP.

H4c. The WOSP evaluation was more accurate.

H4d. The WOSP evaluation was more complete.

C. AHP Tool

We used a criterion based method of system evaluation to compare the WOSP and TAM frameworks. The Analytic Hierarchy Process (AHP) method [79, 80] 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 sub-criteria used to judge the decision alternatives, which are below that again (Figure 5). Pair-wise comparison using decision maker judgments is used to develop weights for the criteria, sub-criteria, 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 [80]. Although the AHP method has been applied to software selection previously [16, 75-77], 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 multi-criterion decision hierarchy (Figure 4) that allowed AHP comparisons.

V. METHOD

The research strategy was to let subjects use the two frameworks to rank software applications, then measure the differences. We used a two-way randomised repeated measures 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 behavioural intention to use, rather than actual use, as prior studies strongly link intention and behaviour [35].

A. Software evaluated

The modern web browser is a powerful, feature-rich application, 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, email, multi-media 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:

- Internet Explorer (current version).
- Mozilla Firefox
- MYIE2

B. Task

Subjects were told 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 study, all communication was by email. Assessing eight WOSP dimensions across three applications seemed about the limit of subject capability, so this procedure could not be used, say, to assess 30 browsers. The process involved four phases:

Phase I. Introduction. Subjects agreed to participate, signed the consent, answered a simple questionnaire, and were asked not to discuss the study with others.

Phase II. Criterion Understanding. Subjects were given the criteria definitions and explanatory statements (Appendix A) and asked to rate them. This phase helped subjects to understand the different criteria.

Phase III. 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 re-submit until consistent responses were achieved. After using both methods subjects were asked which they preferred.

Phase IV. Outcome Evaluation. The AHP responses were processed and the results given to each subject, who then completed a questionnaire on each method's outcome.

D. Criterion understanding

Phase II helped subjects to 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, then had to respond on a 1–5 Likert scale (Strongly Disagree to Strongly Agree) if each

Table 2. 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

statement was:

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 hours per 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 II subjects made 41 AHP comparisons, and repeated some if they were

inconsistent. Phases II and III took 1-2 hours 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 subjects remaining 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 the *outcome attitudes*. The phase III AHP evaluation produced the application rankings and criterion weights. Phase III 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 measure in phase IV, after subjects saw their evaluation outcome. *Outcome satisfaction* was a multi-item scale validated for generic multi-criterion decision processes [81]. It was the response average of:

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 using 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:

1. *Confidence.* How confident are you that the method's data will give the best software choice? (Very Confident to Very Unsure)
2. *Accurate evaluation.* How accurate is the method in testing software performance? (Very Accurate to Very Inaccurate)
3. *Complete evaluation.* The method includes all the dimensions I need in order to evaluate software. (Highly Agree to Highly Disagree)

VI. RESULTS

A. Application Rankings

The TAM and WOSP evaluations produced browser ranks of 1st, 2nd or 3rd. Half the subjects ranked one or more browsers differently using TAM vs WOSP criteria, and 27% of the browser ranks changed. To see if the frameworks performed differently, we took the

Table 3. Performance goals decision matrix and weights

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

Table 4. Criteria correlations

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

Table 5. WOSP & TAM Decision Matrix

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

CR = 0

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 1st, 2nd and 3rd and WOSP rated them 3rd, 2nd and 1st, 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 vs

TAM rank totals was significant ($p=0.013^*$), suggesting 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% [82]. No-one chose MyIE2 as their 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 two), suggesting that WOSP has more evaluation discrimination.

B. Criterion Weights

Table 2 shows the aggregated system element weights (CR = 0.005) and Table 3 the criterion weights (CR = 0.000). The most valued element was receptor (0.32), followed by effector (0.26), boundary (0.21) and internal structure (0.21).

The most valued criteria 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

Criteria correlations

Most of the WOSP criteria correlations (Table 4) were low, and none were significant. The highest correlation was -0.44 (functionality by security). This suggests the WOSP model has relatively modular dimensions. Some constructs assumed connected in the research literature were easily distinguished by our users, e.g.

1. *Privacy* (keeping secrets) vs *Security* (protection from attack): correlation +0.01
2. *Extendibility* (plug-in/import compatibility) vs *Connectivity* (download ability): correlation – 0.29
3. *Usability* (reducing effort) vs *Reliability* (maintaining operations): correlation –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 [83], 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 (Table 5).

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 6 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:

H4a. More satisfied with the WOSP evaluation outcome- supported

H4b. More confident in their choice using WOSP- supported

H4c. The WOSP evaluation was more accurate-

Table 6. TAM vs WOSP Attitudes

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

supported

H4d. 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 study include the application, the subject number, AHP and WOSP.

Application Context

While the WOSP model applies to any socio-technical system, 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 [84]. WOSP criteria weights should vary with the situation, e.g. a fast-food chain, whose staff change 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 [85].

Subject Number

AHP subjects present higher data quality than many other methods, e.g. compare 100 subjects “ticking the boxes” of a questionnaire in say 30 minutes, with 20 people completing an AHP process at 4 hours per person. The twenty AHP subjects represent more data creation effort (80 hours) than 100 questionnaire subjects (50 hours). Since 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.

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, though they often prefer to directly estimate alternatives against criteria [87], which estimates are quicker, but cannot be consistency checked, reducing quality. Despite such limitations, AHP seems a useful tool in multi-criterion comparisons.

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 (ROI), cost-benefit total, return on management and information economics. Multi-objective and multi-criterion (MOMC) methods, value analysis (VA) methods, and critical success factor (CSF) methods, each have advantages and limitations [88]. Combining such established methods with the WOSP model could help managers make better IS/IT evaluations.

B. Implications

Risk vs Opportunity

In this study, 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 spy-ware.

Extend TAM with WOSP

Subjects evaluating socio-technical 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” [86]. This could also explain why IS “killer” applications (like email, 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 and women more on ease of use [19]). If men are in general more opportunity driven, they may favour the WOSP opportunity criteria (functionality, connectivity, flexibility, extendibility), while if women are more risk averse, they may favour the WOSP criteria usability, security, reliability and privacy.

Privacy – a Sleeper Social Requirement

The importance of privacy in our subject’s ratings, most of whom were young, suggests this is a “sleeper” social requirement. The days when Sun Microsystems CEO could say: “*You have zero privacy anyway. Get over it.*” [89] seem over. The quote illustrates the view of technical determinism – that online technical reality 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 “socio-physical” system, just as one defines a socio-technical system. 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 three foot room everyone is three foot 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 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.

A 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 hardware requirements are also met). Similarly social synergy offers enormous productivity increases, given social requirements like privacy and fairness are met. The conclusion is that online prosperity will increase as human and social requirements drive technical design [90].

If one graphs the criteria functionality and usability, their “efficient frontier” is a line that defines the best one can achieve of one criterion for a given value of the other [91]. To design a system that is both functional and usable requires a “synthesis of form”, that reconciles the different requirements [92]. The efficient frontier concept implies 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 two dimensional 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 multi-dimensional space, perhaps why software design is as much an art as a science.

The WOSP model’s implications can be summarized:

1. *System performance is multi-dimensional* - no single design criterion describes all performance requirements, e.g. functionality, usability, reliability, flexibility, security, extendibility, connectivity or privacy can all be critical.
2. *System performance depends on the environment* – a design criterion weights (0-100%) change with the environment, e.g. security has more weight in a threat environment.
3. *Improving one performance criteria 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 armoured vehicles and fast unarmoured ones both perform well in different ways.
5. *Innovations can expand the efficient frontier*, by reconciling criterion conflicts, e.g. light-weight armour that allows fast and armoured 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. *Firstly*, 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.*” [94, p58]. 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*. *Secondly*, the WOSP performance concept can work at different IS levels, and so 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 in modern software [93], suggests this view is not unreasonable. The WOSP advantages are:

1. It offers a useful check-list 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 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:

1. It currently lacks a time dimension, and so must be used with a cost estimate method.
2. Its criteria are general, and must be specifically defined for each application (see Appendix A).
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 Web of System Performance extension to TAM will help designers, users and managers better design operate and evaluate modern socio-technical software.

REFERENCES

1. J.C. Westland and T.H.K. Clark, *Global Electronic Commerce*. Location: MIT Press, 2000, Cambridge, MA.
2. P.B. Seddon, V. Graesser, and L.P. Willcocks, "Measuring organizational IS effectiveness an overview and update of senior management perspectives". *ACM SIGMIS Database*, **33**, 2002.
3. S. Taylor and P.A. Todd, "Assessing IT usage: the role of prior experience". *MIS Quarterly*, **19** (4), p.561-570, 1995.
4. E. Brynjolfsson, "The productivity paradox of information technology". *Communications of the ACM*, **36**, 1993.
5. M.Y. Ivory and M.A. Hearst, "The state of the art in automating usability evaluation of user interfaces". *ACM Computing Surveys*, **33** (4), p.470 - 516, 2001.
6. W.J. Orlikowski and D.G. Gash, "Technological frames: Making sense of information technology in organizations". *ACM Transactions on Information Systems*, **12** (2), p.174-207, 1994.
7. J.R. Ralston, *Voltaire's Bastards*. Location: Penguin Books, 1993, New York.
8. B.N. Reeves and A.C. Lemke, "The problem as a moving target in cooperative system design". 1991 HCI Consortium Workshop, (Jan), 1991.
9. J. Grudin, *Human Factors, CHI and MIS*, in *Advances in Management Information Systems*, P. Zhang and D. Galletta, Editors. 2005, M.E. Sharpe.
10. M. Fishbein and I. Ajzen, *Belief, Attitude, Intention and Behaviour: An introduction to theory and research*. Location: Addison-Wesley, 1975, Reading.
11. F.D. Davis and V. Venkatesh, "A critical assessment of potential measurement biases in the technology acceptance model: three experiments". *Int. Journal of Human-Computer Studies*, **45**, p.19-45, 1996.
12. F.D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology". *Management Information Systems Quarterly*, **Sep**, 1989.
13. D. Hix and R.S. Schulman, "Human-computer interface development tools: A methodology for their evaluation". *Communications of the ACM*, **34** (3), p.75-87, 1991.
14. J. Lin, Lu, H., "Towards an understanding of the behavioral intention to use a web site". *International Journal of Information Management*, **20**, p.197-208, 2000.
15. L. Vijayarathy, "Predicting consumer intentions to use on-line shopping: The case for an augmented technology acceptance model". *Information & Management*, **41**, p.747-762, 2004.
16. V. Lai and H. Li, "Technology acceptance model for internet banking: an invariance analysis". *Information & Management*, **42**, p.373-386, 2005.
17. H. Heijden, "Factors influencing the usage of websites: the case of a generic portal in the The Netherlands." *Information & Management*, **40**, p.541-549, 2003.
18. P. Legris, J. Ingham, and P. Collette, "Why do people use information technology? A critical review of the technology acceptance model". *Information & Management*, **40**, p.191-204, 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 Quarterly*, **24** (1), p.115-139, 2000.
20. I. Brown, "Individual and technological factors affecting perceived ease of use of web-based learning technologies in a developing country". *The Electronic Journal on Information Systems in Developing Countries*, **9** (5), p.1-15, 2002.
21. V. Venkatesh, "Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model." *Information Systems Research*, **11** (4), p.342-365, 2000.
22. J. Moon and Y. Kim, "Extending the TAM for a World-Wide-Web context". *Information & Management*, **38**, p.217-230, 2001.
23. C. Ong, J. Lai, and Y. Wang, "Factors affecting engineer's acceptance of asynchronous e-learning systems in high-tech companies". *Information & Management*, **41**, p.795-804, 2004.
24. H. Shih, "An empirical study on predicting user acceptance of e-shopping on the Web". *Information & Management*, **41**, p.351-368, 2004.
25. J. Yu, et al., "Extending the TAM for a t-commerce". *Information & Management*, **In Press**, 2005.
26. I. Benbasat and H. Barki, *Quo Vadis TAM*. 2006: Correspondence to the first Author
27. L. Festinger, *A Theory of Cognitive Dissonance*. Location: Stanford University Press, 1957.
28. V. Venkatesh and M.G. Morris, "User acceptance of information technology: toward a unified view". *MIS Quarterly*, **27** (3), p.425-478, 2003.
29. S.E. Asch, *Social Psychology*. Location: NY Prentice Hall, 1952.
30. R.S. Crutchfield, "Conformity and character". *American Psychologist*, **10**, p.191-198, 1955.
31. M. Lea and R. Spears, "Computer-mediated communication, de-individuation and group decision making". *International Journal of Man-Machine Studies*, **34**, p.283-301, 1991.
32. T. Postmes, *Social Influence in Computer-Mediated Groups*. Location: University of Amsterdam, 1997, Amsterdam: 165.
33. B. Whitworth, B. Gallupe, and R. McQueen, "Generating agreement in computer-mediated groups". *Small Group Research*, **32** (5), p.621-661, 2001.
34. B. Whitworth, "Polite Computing". *Behaviour & Information Technology*, (5, September, <http://brianwhitworth.com/polite.rtf>), p.353 - 363, 2005.
35. P.J. Hu, et al., "Examining the technology acceptance model using physician acceptance of telemedicine". *Journal of Management Information Systems*, **16** (2), p.91-112, 1999.
36. I. Sommerville, *Software Engineering*. 7th ed., 2004.
37. D. Compeau, C.A. Higgins, and S. Huff, "Social cognitive theory and individual reactions to computing technology: A longitudinal study". *MIS Quarterly*, **23** (2), p.145-158, 1999.
38. G. Moore and I. Benbasat, "Development of an instrument to measure the perceptions of adopting an information technology innovation". *Information Systems Research*, **2** (3), p.192-222, 1991.
39. K.E. Boulding, "General systems theory - the skeleton of a science". *Management Science*, **2** (3), p.197-208, 1956.
40. J. Grudin. "The Computer Reaches Out: The historical continuity of user interface design." in *Proceedings of CHI '90, ACM SIGCHI Conference*. 1990. Seattle, Wash., USA: ACM.
41. K. Kuutti, *Activity Theory as a Potential Framework for Human Computer Interaction Research*, in *Context and Consciousness: Activity Theory and Human-Computer Interaction*, B.A. Nardi, Editor. 1996, The MIT Press: Cambridge, Massachusetts.
42. S. Alter, "A general, yet useful theory of information systems". *Comm. of the AIS*, **1** (March), p.13-60, 1999.
43. C.E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Location: University of Illinois Press, 1949, Urbana.

44. B. Whitworth, Social-technical systems in Encyclopaedia of Human Computer Interaction, C. Ghaoui, Editor. 2006, Idea Group Reference: London, p. 533-541.
49. B. Whitworth, B. Gallupe, and R.J. McQueen, "A cognitive three process model of computer-mediated groups: Theoretical foundations for groupware design". *Group Decision and Negotiation*, **9** (5), p.431-456, 2000.
50. F. Losavio, et al., "Designing quality architecture: incorporating ISO standards into the unified process". *Information Systems Management*, **21** (1), p.27 - 44, 2004.
51. T. Berners-Lee, *Weaving The Web: The original design and ultimate destiny of the world wide web*. Location: Harper-Collins, 2000, New York.
52. A. Gargaro, et al., "The Power of Standards." *Communications of the ACM*, **36** (8), p.11-12, 1993.
53. M. De Simone and R. Kazman. "Software Architectural Analysis: An Experience Report". in *Proceedings of CASCON'95*. 1995. Toronto, ON.
54. K. Knoll and S.L. Jarvenpaa. "Information technology alignment or "fit" in highly turbulent environments: The concept of flexibility". in *SIGCPR 94*. 1994. Alexandria, Virginia, USA: ACM.
55. P. Regan, *Legislating privacy, technology, social values and public policy*. Location: University of North Carolina Press, 1995, Chapel Hill, NC.
56. J.C. Laprie and A. Costes. "Dependability: A unifying concept for reliable computing". in *Proc 12th International Symposium on Fault-Tolerant Computing*. 1982. FTCS-12: IEEE.
57. OECD, *Guidelines for the Security of Information Systems*. Location: OECD, 1996: 50.
58. E. Jonsson. "An integrated framework for security and dependability". in *NSPW*. 1998. Charlottesville, VA, USA: ACM.
59. G. Gediga, K. Hamborg, and I. Dunsch, "The IsoMetrics usability inventory: an operationalization of ISO9241-10 supporting summative and formative evaluation of software systems". *Behaviour & Information Technology*, **18** (3), p.151-164, 1999.
60. B. Nuseibeh and S. Easterbrook, *Requirements Engineering: A roadmap*, in *The Future of Software Engineering*, A. Finkelstein, Editor. 2000, ACM Press: Limerick, Ireland.
61. J.G. Walls, G.R. Widmeyer, and O.A. El Sawy, "Building and information system design theory for vigilant EIS". *Information Systems Research*, **3** (1), p.36-59, 1992.
62. A.R. Hevner, S.T. March, and J. Park, "Design Science in Information Systems Research". *MIS Quarterly*, **28** (1), p.75-105, 2004.
63. L.v. Bertalanffy, *General System Theory*. Location: George Braziller Inc, 1968, New York.
64. L. Chung, et al., *Non-functional requirements in Software Engineering*. Location: Kluwer Academic, 1999, Boston.
65. B. Whitworth, J. Fjermestad, and E. Mahinda, "The Web of System Performance: A multi-goal model of information system performance". *Communications of the ACM*, **49**, May (5), p.93-99, 2006.
66. B. Whitworth and M. Zaic, "The WOSP Model: Balanced Information System Design and Evaluation". *Communications of the Association for Information Systems*, **12**, p.258-282, 2003.
67. K. Patten, et al. "Leading IT Flexibility: Anticipation, Agility and Adaptability". in *Americas Conference on Information Systems*. 2005. August 11-14, Omaha, Nebraska, USA: AIS.
68. O.A. El Sawy and B. Nanus, "Towards the design of robust information systems". *Journal of Management Information Systems*, **5** (4), p.33-54, 1989.
69. B.A. Nixon. "Managing performance requirements for information systems". in *Proceedings of the first international workshop on Software and performance*. 1998. Santa Fe, New Mexico, United States: ACM Press New York, NY, USA.
70. N.S. Rosa, G.R.R. Justo, and P.R.F. Cunha. "A Framework for Building Non-Functional Software Architectures". in *SAC*. 2001. Las Vegas, NV: ACM.
71. L.M. Cysneiros and J. Leita. "Non-functional requirements: From elicitation to modelling languages". in *ICSE*. 2002. Orlando, Florida: ACM.
72. J.K. Pinto, "Project Management 2002". *Research Technology Management*, **45** (2), p.22-37, 2002.
73. D.M. Kienzle and W.A. Wulf. "A practical approach to security assessment". in *New Security Paradigms Workshop*. 1997. Langdale, Cumbria, U.K.: ACM.
74. N.S. Borenstein, "Power, ease of use and cooperative work in a practical multimedia message system". *International Journal of Man-Machine Studies*, **34** (2), p.229-259, 1990.
75. B. Whitworth and E. Whitworth, "Reducing spam by closing the social-technical gap". *Computer*, (October), p.38-45, 2004.
76. A. Moreira, J. Araujo, and I. Brita. "Crosscutting quality attributes for requirements engineering". in *SEKE*. 2002. Ischia, Italy: ACM.
77. W.H. De Lone and E.R. McLean, "Information Systems Success: The Quest for the Dependent Variable". *Information Systems Research*, **3** (1), p.60-95, 1992.
78. I. Benbasat and R.W. Zmud, "The identity crisis within the IS discipline: Defining and communicating the discipline's core properties". *MIS Quarterly*, **27** (2), p.183-194, 2003.
79. T.L. Saaty, "A scaling method for Priorities in Hierarchical Structures". *Journal of Mathematical Psychology*, **15** (3), p.234-281, 1977.
80. T.L. Saaty, *The Analytic Hierarchy Process*. Location: McGraw Hill, 1980, New York.
81. Q. Zhang, J. Chen, and P. Chong, "Decision consolidation: Criteria weight determination using multiple preference formats". *Decision Support Systems*, **38**, p.247-258, 2004.
82. B. Whitworth and R. Felton, "Measuring disagreement in groups facing limited choice problems". *THE DATABASE for Advances in Information Systems*, **30** (3 & 4), p.22-33, 1999.
83. T.L. Saaty and L.G. Vargas, *Models, Methods, Concepts & Applications of the Analytic Hierarchy Process*. Location: Kluwer Academic Publishers, 2001, Boston.
84. W.G. Chismar and S. Wiley-Patton. "Does the extended technology acceptance model apply to physicians?" in *Proceedings of the 36th Hawaii International Conference on System Sciences*. 2002. Hawaii: IEEE.
85. E. Mahinda and B. Whitworth. "The Web of System Performance: Extending the TAM Model". in *AMCIS (Americas Conference on Information Systems) August 11-14*. 2005. Omaha, Nebraska, USA.
86. E. Tenner, (). *Why Things Bite Back*. Vintage Books, Random House, 1997 New York.
87. C. Sylla and B. Arinze, "A method for quality precoordination in a quality assurance information system". *IEEE Transactions on Engineering Management*, **38** (3), p.245-256, 1991.
88. C. Sylla and J. Wen, "A conceptual framework for evaluation of information technology investments". *International Journal of Technology Management*, **24** (2/3), p.236-261, 2002.
89. S. Manes, "Private Lives? Not Ours!" *PC World*, **June**, p.312, 2000.
90. B. Whitworth and A. deMoor, "Legitimate by design: Towards trusted virtual community environments". *Behaviour & Information Technology*, **22** (1), p.31-51, 2003.
91. R.L. Keeney and H. Raiffa, *Decisions with multiple objectives: Preferences and Value Tradeoffs*. Wiley, 1976.
92. C. Alexander, *Notes on the Synthesis of Form*. Location: Harvard University Press, 1964, Cambridge, Ma.

93. J. S. David, W. E. McCarthy & B. S. Sommer, "Agility - The key to survival of the fittest" Communications of the ACM, 46(5), 65-69, 2003.

94. R. Koch, "Flexibility or Performance?" Scientific American, August, p58, 2005

APPENDIX A. TAM/WOSP CRITERIA AND STATEMENTS

TAM

The TAM criteria and grounding statements were [18]:

1. *Usefulness (PU): That the software is useful*

- Using the browser increases my productivity
- Using the browser increases my job performance
- Using the browser enhances my effectiveness on the job

2. *Ease of use (PEOU): That the software is easy to use*

- Learning to operate this browser is easy for me
- I find it easy to get the this browser to do what I want to do
- This browser is not rigid and inflexible to interact with

WOSP

The WOSP selection criteria were first broken down by the 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 in 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 sub-criteria 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*

- It works with all third party multi-media tools, like real-media player and flash.
- It follows all World Wide Web source code and data standards, e.g. unicode.
- 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*

- When a file is downloaded to the hard drive, it is checked for viruses before use.
- It can detect and prevent spyware from installing.
- It can detect and prevent popup ads.

3. *Reliability: The ability to continue working despite errors/problems, or to quickly recover from failure*

- It never breaks down or "hangs" (fails to respond), even if I use it for a long time.
- If one part of the browser fails, like a plug-in, the entire browser does not crash.
- Even if I multi-task, and do many things at once, it still works well.

4. *Flexibility: The ability to change itself to fit different situations*

- It runs on all our computers and operating systems.
- It is easily changed to fit disability needs e.g. larger text or graphics for those with poor vision.

- It has a preferences "control panel" to change browser settings.

5. *Functionality: The ability to move to web sites and display their information*

- The Favourites list lets me jump directly to my favourite sites
- This browser gets me where I want to go quickly
- The browser has everything I need to search, navigate and display the Internet.

6. *Usability: The ability to be used easily*

- The user interface is consistent and easy to learn.
- I did not need training to use it the first time.
- I accomplish my tasks easier and quicker with this browser.

7. *Connectivity: The ability to exchange information with other systems*

- When downloading it gives useful information, like the estimated time to complete the download.
- 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.
- It gives access to other ways of communicating, like Telnet, Ftp, email and chat.

8. *Privacy: The ability to limit unwanted information disclosure*

- Any sensitive information I give the browser, like logon passwords, is encrypted, so others can't see it.
- Password information always shows as asterisks, so others cannot look over my shoulder to see them
- It stops web sites from getting my name or email from my computer's data.

Brian Whitworth, is a Senior Lecturer at Massey University (Albany), New Zealand. His psychology Masters thesis was on the implications of split-brain research for the concept of "Self". After time as the New Zealand Army Senior Psychologist, he worked as a defence computer systems analyst and in operational simulations. After "retiring", he designed and wrote the groupware used in his doctorate on generating computer-mediated agreement. He has published in Small Group Research, Group Decision & Negotiation, THE DATABASE for Advances in IS, Communications of the AIS, IEEE Computer, Behaviour and Information Technology (BIT) and Communications of the ACM. His main area of interest is the design and evaluation of socio-technical systems. See <http://brianwhitworth.com/papers.html>



Victor A. Bañuls is currently an assistant professor of Management Information Systems at Pablo de Olavide University at Seville (Spain). He holds a B.S. in Business Administration from University of Huelva and a professional M.S. in Computer Systems. Before his current position, Dr.

Bañuls has been working as information systems consultant and has participated in several international research projects. He also served as a Visiting Research Scholar at the Information Systems Department of the New Jersey Institute of Technology, New Jersey (U.S). His research has been published in journals such as Technological Forecasting and Social Change, Futures, Technovation and others. At present, his main research interests are information systems assessment and technology foresight methodologies.

Cheickna Sylla is an Associate Professor of Management Science and Decision Science at the School of Management, New Jersey Institute of Technology (NJIT), Newark, New Jersey. He holds a Ph.D. in Industrial Engineering from SUNY at Buffalo, Buffalo, New York. 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. His research publications, over 60, appear in IEEE Transactions in Systems, Man and Cybernetics, IEEE Transactions on Engineering Management, European Journal of Operations Research, Human Factors, Decision Sciences, Computers and Industrial Engineering, Control Engineering Practice, Cost Management, International Journal of Management and Decision Making (IJMDM), International Journal of Technology Management (IJITM), the International Journal of Information and Technology Management, and numerous conference proceedings.