Claims Reuse for Notification Systems Design: LINK-UP Vision and IRC Equations

C. M. Chewar and D. Scott McCrickard

Center for Human Computer Interaction and Department of Computer Science
Virginia Polytechnic Institute and State University
Blacksburg, VA 24061-0106

{cchewar, mccricks}@cs.vt.edu

Abstract

Extending previous work on the concept of claims reuse, an approach for cataloging and reusing design knowledge in human-computer interaction, we introduce a vision for a system, LINK-UP. The system is intended to parallel a usability engineering process that involves claims analysis. While we initially target notification system design support, we believe that the general method is extensible to other design concerns. A key aspect of the LINK-UP system is its iterative assessment of critical parameters—essential target values that describe anticipated user goals. In notification systems design, three critical parameters are interruption, reaction, and comprehension, referred to as *IRC*. While the parameter values represent abstract concepts, a pivotal challenge in the development of LINK-UP is determining methods for consistent and accurate parameter specification. To this end, we introduce equations for calculating user's model IRC parameters, either from analytical or empirical data. Presented here are details of variable justification and equation behavior. Future work will assess consistency and accuracy of artifact classifications using the equations.

1 LINK-UP Overview

The LINK-UP system (Leveraging Integrated Notification Knowledge through Usability Parameters), which will be developed and validated in ongoing and future research, demonstrates an extended approach to claims analysis in usability engineering. *Claims* are statements about the effects of a designed feature on a user in the context of a usage scenario [8]. An example of a simple claim could be:

Use of **tickering text-based animation** to display news headlines in a small desktop window:
Preserves user focus on a primary task, while allowing long-term awareness
BUT is not suitable for rapid recognition of and reaction to urgent information.

Claims are grounded by empirical testing or observation, so a designer of a notification system may compare this claim with claims related to use of in-place animation techniques, such as fading and blasting. To simplify the process, we can abstract a claim and focus on *critical parameters* relating information presentation to effects on information processing—user's interruption, reaction, and comprehension (or *IRC*):

Tickering text-based animation ∈ {*low* interruption, *low* reaction, *moderate* comprehension}

LINK-UP is specifically for design and evaluation of notification systems, but it can be generalized to other classes of system through future efforts. As designers proceed through a design cycle, they continuously question the values of targeted and actual critical parameters for important interface aspects. Claims are stored a design knowledge repository, and are accessed and modified at several points with interactive system tools.

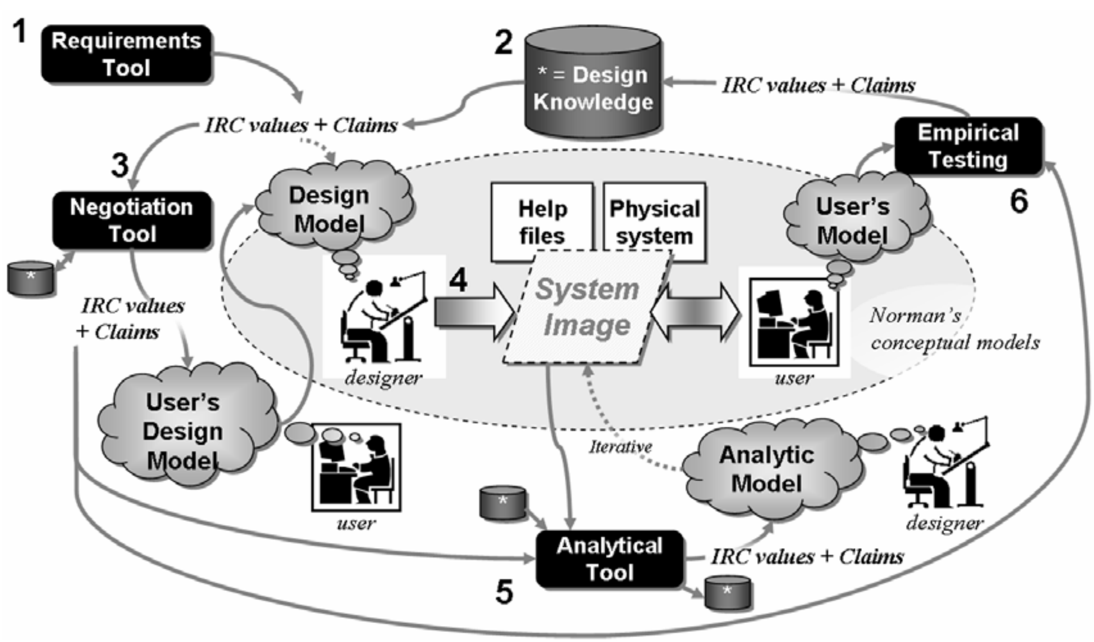


Figure 1. *General architecture of LINK-UP, the proposed usability engineering process and system. The light grey region in the center depicts Norman's conceptual models [7], which are extended through proposed work. Numbers refer to steps though the process, and are referenced and explained throughout the following sections*

1.1 Requirements Analysis—Understanding the Design Model

The first concern in a usability engineering process, and the first step in using the LINK-UP system ("1" in Figure 1), is gathering and analyzing user requirement that will drive interface design. Just as with other interface design, requirements analysis for notification systems implies understanding tasks, information characteristics, user background, and other aspects of the situation—forming the *design model*. Although many dimensions of successful dual-task design have been recognized in research (see [11]), notification systems designers need convenient access to these considerations. These variables form the critical parameter levels of desirable user interruption, reaction, and comprehension (IRC values).

To facilitate this, we will create a web-based questioning system that probes requirements relating to the critical parameters. Questions must be easy to understand and guide reasoning about notification tasks and usage factors. The underlying algorithm that converts responses to IRC values (transparent to the designer) must be accurate and consistent for a wide variety of design models. The process will indirectly educate designers about dual-task design concerns, providing links to detailed research summaries. We have already completed preliminary work on this requirements tool and achieved encouraging results though user testing, although further development, testing, and documentation is needed. Methods that guide development and validation for accuracy and consistency include expert walkthroughs with a variety of systems and lab-based testing of novice designers that are given general design problem specifications. Once targeted IRC values are calculated, they serve as an index for accessing claims in a design knowledge repository. Other indices are used as well, such as the generic tasks that the system will support (e.g., monitoring or alerting) and design choices (e.g., use of color or animation). Using the LINK-UP system, designers will be able to search for influential and reusable claims and store them in a manner similar to the Internet shopping cart metaphor used on e-commerce sites ("2" in Figure 1).

1.2 Negotiating the User's Design Model with Claims

The claims collected in step 2 will assist designers in reasoning about activity, information, and interaction design choices (scenario-based design phases, see [8]) related to the developing notification system. However, to assist in participatory design efforts and validate the design model IRC values, the LINK-UP system will provide a tool for designers that produces an interactive claims-review session, for use with potential users ("3" in Figure 1). Designers will be able to present prototypical usage scenarios to the user, who will then assess the claims (and underlying, transparent IRC values). Users will be able to accept or reject claims according to their needs, forming the *user's design model*—a conception of the system effects that can be gleaned through the IRC values associated with the final claims set. In turn, this model should inform the design model and allow production of the physical system ("4" in Figure 1). It is anticipated that designer-user claims negotiation will be an iterative process involving multiple users. Therefore, the supporting tool will be a robust interface. To formatively test this system, we will run a lab-based experiment with several designer-user dyads to negotiate requirements for multiple scenarios and systems. Case study development efforts will provide summative validation.

1.3 Forming an Analytical Model through Expert Evaluation

Once a system image is available, the LINKUP system will support analytical (expert) evaluation, with the hope that most usability problems can be caught early in the development process and without requiring costly user evaluation. This stage will require adaptation of an analytical evaluation method, such as a cognitive walkthrough or use of heuristics, to notification design. However, LINK-UP will integrate the analytical method with the larger system, providing a tool ("5" in Figure 1) that facilitates execution of the method, recording of results, and estimation of the actual IRC values, or the *analytical model*. In this step, the claims set corresponding IRC values are assessed in light of the physical product, guided by the analytical method. Designers will be able to gauge whether targeted critical parameters will be achieved in the design, receiving automated support to pinpoint specific design problems. Since claims are very much hypotheses before this point and this step provides the first opportunity for hypothesis testing, the tool will allow designers to revise information associated with the claims.

1.4 Confirming Usability with Empirical Testing

Similar to the previous step, the next tool within the LINK-UP system facilitates the execution and results analysis for an empirical user testing session ("5" in Figure 1). Again, the system will use the original set of claims to adapt a general instrument for collecting usage data. Based on users' qualitative feedback and quantitative performance, actual IRC values will be determined to characterize the *user's model* and effectiveness of the claims. While the step allows formative and summative testing of the designed interface, it will also generate new knowledge related to new and existing claims. The key function of the tool assists the designer in comparing actual with intended efforts, informing the next design iteration. However, since the process also calculates the *user's model* IRC value, the system can be compared with other notification systems based on classification in the IRC framework. While preliminary work focuses on creation and testing both the analytical and empirical testing methods, the functions of both tools will require in-depth testing with system walkthroughs, expert review, and case analysis.

1.5 Integrated Design Knowledge Reuse

A main advantage of the LINK-UP system is that it provides continuous and integrated access to the design knowledge repository, facilitating knowledge reuse. The design knowledge repository will build directly on a working prototype system that is based on Sutcliffe's Domain Theory [9]. Through access to the claims database, designers will be able to build from and test previous design results. They will also be able to contribute to a growing body of knowledge. To enable these features in a manner that preserves content quality and user trust, the system will also include accounts and profiles of designers and expert administrators. Expert access to the claims database will allow full claims administration, association of claims with related theories, example systems, and design artifacts, and other meta-analysis features. We will also explore the use of other knowledge management features, such as a claims entry, editing, rating, and commenting features for designers. Results achieved with this aspect of the project will be interest to the wider knowledge reuse community.

1.6 System Summary

To summarize, the LINK-UP system will provide a web-based interface to guide the usability engineering process for a notification system. Designers will be able to interact with five major design support tools (to include support for requirements analysis and negotiation, analytical and empirical testing, and design knowledge access), saving and building on progressive session results throughout the process. A set of claims (serving as design hypotheses) and associated critical parameters (serving as engineering targets and results) guide design progress, within a single design and through a meta-analysis of several systems. The design knowledge repository will grow and improve through use, becoming a living record of notification systems research. The LINK-UP system is a major focus of current research group efforts, and should be publicly available within five years. While much of the preliminary work has been completed, each tool requires additional development, integration within the larger system, and formative evaluation.

2 Calculating IRC

This material provides technical description of the critical parameters for notification systems, *interruption, reaction*, and *comprehension* (IRC). As these are general and abstract concepts, the remainder of the document describes a concrete procedure for assessing component properties of each parameter through user experience. For more information on IRC, to include a justification of parameter selection and intended usage, see [5, 6].

2.1 Variables

	Description	Assessment Technique	
		Analytical/Subjective	**Empirical/Objective**
COI	cost of interruption	*Given the nature and importance of the user's primary task at the receipt of the notification, how **costly** would an interruption be?* {extremely = 1; very = .75; moderately = .5; not very = .25; not at all = 0}	Interruption Workbench [4] output; P(High) is weighed at 1, P(Med) = .5, P(Low)=0
s	primary task sustainment	*Compared to the primary task performance before the notification delivery, how much does the primary task performance **reduce** when the*	Ptask performance while multitasking divided by ptask performance as a solo-task

		notification is present? {not at all = 1; less than half = .75; about half = .5; more than half = .25; completely stops = 0}	[10]
h	hit rate	*How often will users actually notice important changes in the notification, as opposed to not noticing them?* {always = 1; more than half = .75; about half = .5; less than half = .25; never = .0001}	As in Signal Detection Theory, P(H) divided by total signals [3,11]
t	response time	*In cases where a notification suggests an action for a user to take, how does the user's response time compare to the reasonably desired response time?* {better or as good as expected = 1; slightly slower = .75; about twice as slow as expected = .5; much slower = .25; extremely slow or action never taken; no action ever required = 1}	Determine *actual response time* (a) as the difference between signal presentation and signal response; *expected response time* (e) provided in system specification; $t = e / a$, when $a > e$ (otherwise $t = 1$)
p	perception rate	*When considering the total number of times a user interacts with the notification system, what is the* ***ratio*** *of the interactions in response to an important notification vs. total interactions (including those when no actual notification was being delivered, i.e., user checking on their own or thinking there was a notification)?* {1 to 1 = 1; 2 to 3= .75; 1 to 2 = .5; 1 to 4 = .25; more than 1 to 4 = 0}	As in Signal Detection Theory, P(H) divided by total responses [3,11]
c	base comprehension	*How much of the notification content will the user want to remember <u>and</u> be able to remember several minutes after the notification is delivered?* {all content = 1; more than half = .75; about half = .5; less than half = .25; none at all = 0}	Quiz user on a sample of notification content questions to assess correct interpretation, relationship to goals, and storage in long term memory. Use % correct.
f	projection	*Based on the notification content, how successful will the user be in making projections or predictions about future trends or the long-term state of the system being monitored?* {extremely successful = 1; very successful = .75; somewhat successful = .5; not very successful = .25; not a goal for this system = 0}	Quiz user based on a sample of interpretations that can be projected to predict future states or notification patterns. Use % correct.

2.2 Equations

Three equations are provided to allow combination of variables into the abstract parameters. This is not intended to be a robust, integer-based system. Instead, the equations are intended as a conceptual metaphor, loosely organized as a categorical, interval scale approximation. When considering the validity of the equations, one should think of them as numeric representations of *low, somewhat low, moderate, somewhat high*, and *high* parameter categories. The equations are thought to assist in obtaining more consist selection of these categories while assigning user's model parameter values (an additional algorithm has been developed to assess design model IRC parameters), and numeric representations are used to facilitate searching/indexing operations. While testing of this hypothesis is ongoing and results will be submitted for publication, at this time, preliminary results have been promising. Current efforts implement the analytical/subjective questions in a web-based form, facilitating calculation of parameter values.

Interruption: $\mathbf{I} = 1 - s^{3 \cdot COI}$

Justification

- *I* describes both the appropriateness of an interruption, as well as the actual interruptive effect of the notification artifact (distraction to the primary task). Therefore, "low I" can describe either an artifact that supports attention grading/parallel processing during the performance of an urgent primary task (high sustainment, regardless of COI) or any quality of multitasking performance in a non urgent situation (low COI, regardless of sustainment).

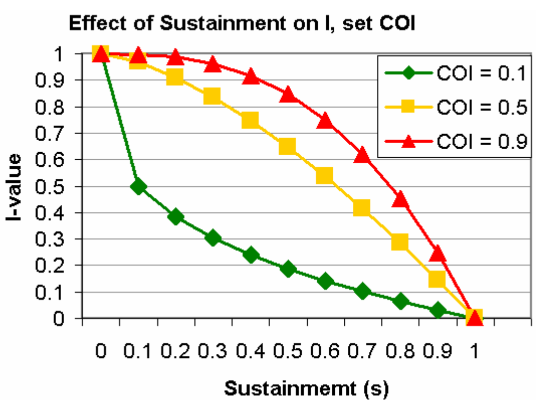
- *Appropriateness of an interruption* is represented by COI (cost of interruption), characterizing the user's willingness to accept an interruption, and thus the urgency of the primary task can be inferred. As established by Horvitz's Interruption Workbench [4], COI describes a total task situation in terms of how much a given user would typically pay in dollars not to be interrupted. The Interruption Workbench records a variety of situation characteristics, such as the specific primary task application, level of ambient noise, recent keystroke and mouse activity, etc) over an extended period of normal user activity. The tool segments the observations into periods in which the task variable combinations are consistent. Users rate each segment, assigning the dollar value they would pay to avoid interruption, allowing Bayesian inference networks to aggregate samples and determine probability distributions for various costs of interruption levels.

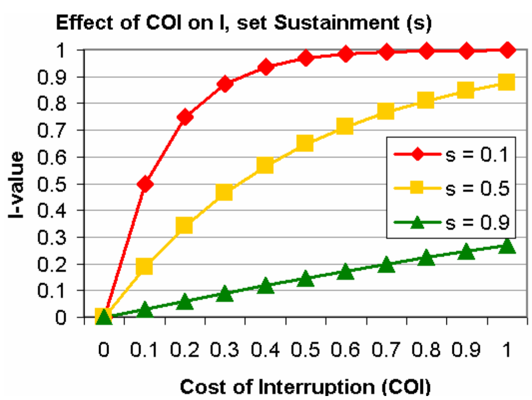
- *Actual interruptive effect* can be gauged by primary task sustainment—a metric used to quantify the change in the primary task performance from solo-task performance to dual-task performance.

- The equation is modeled with an exponential COI to reinforce the importance of this factor, but tripled to ensure a fairly wide range of I-values for a given COI and to produce a moderately high I-value (0.65) when both s and COI = 0.5.

Sample results:

COI	s	**I**	Rationale
0.1	0.9	0.03	Very low cost, very low actual interruption—likely to employ **discrete attention division**
0.1	0.5	0.19	
0.1	0.1	0.50	Classic case of **user initiated interruption**
0.3	0.9	0.09	
0.3	0.5	0.46	
0.3	0.1	0.87	User initiated interruption
0.5	0.9	0.15	Moderately important primary task effectively preserved, combination of graded and discrete attention division
0.5	0.5	0.65	
0.5	0.1	0.97	User initiated interruption possible
0.7	0.9	0.20	
0.7	0.5	0.77	
0.7	0.1	0.99	User initiated interruption possible, but not likely
0.9	0.9	0.25	Urgent primary task effectively preserved, probably through **graded attention division**
0.9	0.5	0.85	
0.9	0.1	0.99	Urgent primary task almost completely interrupted—most likely an extremely valued notification or a design flaw





Reaction: $$\mathbf{R} = \frac{(t \cdot h)^{\frac{1}{3 \cdot COI}}}{2} + \frac{h(0.5 + COI)}{3}$$

Justification

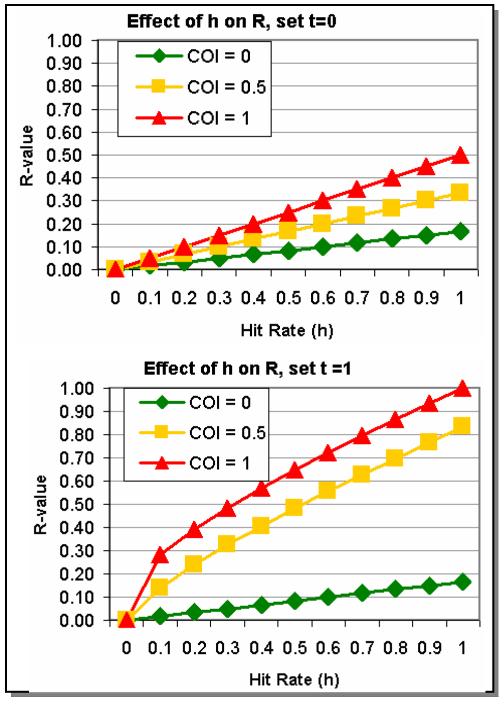
- The R equation consists of two parts, each worth up to an R-value component of 0.5. The first term takes two reaction performance metrics—*hit rate* (*h*) and *response time* (*t*)—and lowers their average according to strength of COI. The second term can add up to half the hit rate to the R-value, depending on the strength of COI. Moderate reaction (R=0.5) is scored when two-thirds of the hit rate and reaction time is achieved with a COI of 0.5. Moderate or high R-values are always obtained when one of the variables is near maximum and the others are at least moderate.

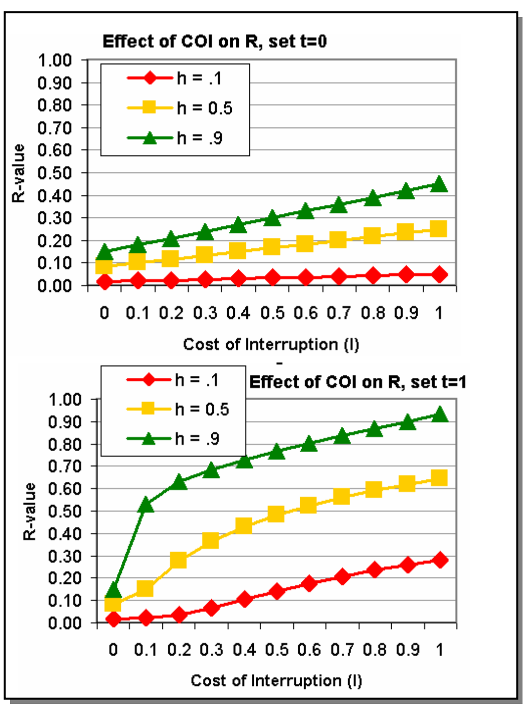
- The equation is also designed so that no more than R=0.5 can be achieved if one of the three variables equals zero. In order to understanding this rationale, one must considering that R is a characterization of an artifact's effectiveness for supporting reaction in a dual-task situation. That is, if the notification system is not attempting to resolve a situation constrained by the attention-utility theme (in which there would generally be at least a moderate value for COI), then the appeal of the artifact for facilitating notification reaction in a dual-task situation is inherently limited and therefore penalized. Both aspects of the reaction performance are also critical—a near-perfect hit rate would not be looked at as effective reaction if the response time is significantly slower than specification. Likewise, an acceptable response time has limited worth in the case that most signals delivered by the notification system are missed.

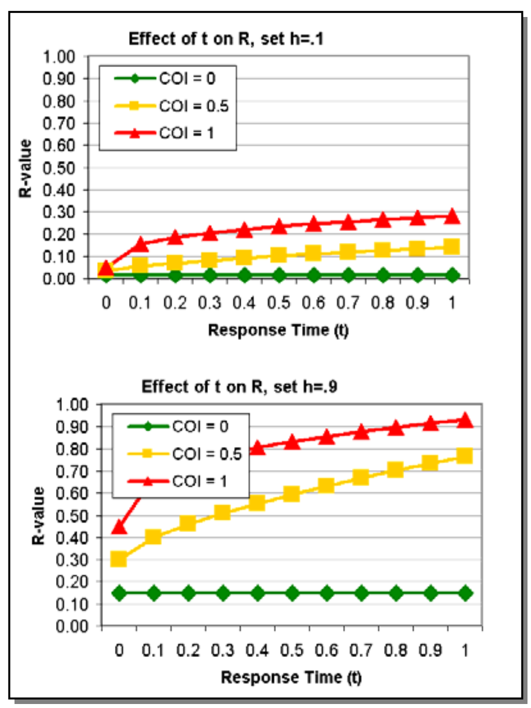
- Another feature of the equation is the prominence of the hit rate. Factoring this variable directly into both terms allows quick growth of R-values as hit rate increases, especially when COI is greater than 0.5. This adds a strong characteristic to R of being a measure of response selection probability. Since the test protocol relies on definition of total number of signals present, evaluators should ensure users are only expected to respond to a realistic number of important notifications. A suggested guideline for setting notification system test specifications is that expected reaction times and all signals can be hit at 100% without the primary task.

Sample results:

h	t	COI	**R**	Rationale
0.1	0	0	0.02	Certainly an **information design problem (perception)**—make notification more prominent
0.1	0	1	0.05	
0.1	0.5	0	0.02	Likely to be employing user initiated interruption or discrete attention division at inappropriate intervals—perhaps use an affordance to prompt monitoring cadence
0.1	0.5	1	0.23	Poor hit rate, but adequate response time in an urgent situation
0.3	0.5	0	0.05	
0.3	0.5	1	0.42	Decent response rate in an urgent situation, but too many signal are missed in this case
0.5	0	0	0.08	
0.5	0	1	0.25	
0.5	0.5	0	0.08	
0.5	0.5	0.5	0.37	Only moderate performance metrics in a moderately urgent situation—better than low R, but not by much
0.5	0.5	1	0.56	
0.7	0.5	0	0.12	
0.7	0.5	0.5	0.48	
0.7	0.5	1	0.70	
0.9	0	0	0.15	Hits most signals, but far too late and in a non-urgent situation. Most likely an **interaction design problem (pre-execution stage)**.
0.9	0	0.5	0.30	
0.9	0	1	0.45	Strong stimulus detection in an urgent situation, but extremely slow reaction time—likely an **interaction design problem (execution stage)**
0.9	0.5	0	0.15	
0.9	0.5	0.5	0.59	
0.9	0.5	1	0.83	
0.9	1	0	0.15	Although this artifact facilitates excellent reaction performance metrics, the non urgency of the dual-task situation prevents a strong R-value. Recommend retesting in a more urgent dual-task environment.
0.9	1	0.5	0.77	
0.9	1	1	0.93	


Effect of h on R, set t=0
COI = 0
COI = 0.5
COI = 1
R-value
Hit Rate (h)
Effect of h on R, set t =1
COI = 0
COI = 0.5
COI = 1
R-value
Hit Rate (h)


Effect of COI on R, set t=0
h = .1
h = 0.5
h = .9
R-value
Cost of Interruption (I)
Effect of COI on R, set t=1
h = .1
h = 0.5
h = .9
R-value
Cost of Interruption (I)


Effect of t on R, set h=.1
COI = 0
COI = 0.5
COI = 1
R-value
Response Time (t)
Effect of t on R, set h=.9
COI = 0
COI = 0.5
COI = 1
R-value
Response Time (t)

Comprehension: $\mathbf{C} = f + \dfrac{(1-f)(p+2c-cp)}{3}$

Justification

- Based on concept of situation awareness, where a user accumulates Perception (of the elements in the system), Comprehension (of the current situation), and then Projection (of future status). Each level is dependent on achieving some part of the preceding level, and represents a progressively higher state of situated awareness [2]. Thinking of *C* as situation awareness brings it in line with a wealth of research in the human factors field. For instance, studies have shown that we can recognize the characteristics of awareness independent of the processes required to maintain it (working and long term memory or attentional state) [1] or the response selections that result from it [11].

- The simplified equation that appears above is difficult to explain, so we revert to the unsimplified version below:

$$\mathbf{C} = \frac{p+(1-p)(c+f(1-c))}{3} + \frac{c+f(1-c)}{3} + \frac{f}{3}$$

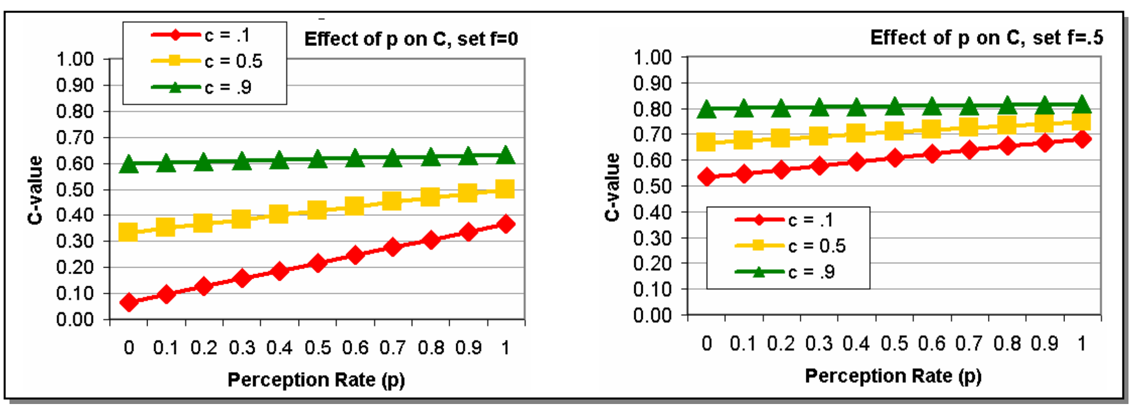
This equation consists of three terms—one for each level of situation awareness. As each level is maximized, the equation ensures that C=0.33, 0.67, and 1 respectively. If a given level is not maximized, achievement in the higher levels provide credit toward the C-value. This scoring scheme relies on an experimental protocol that discourages participant guessing or uncertainty.

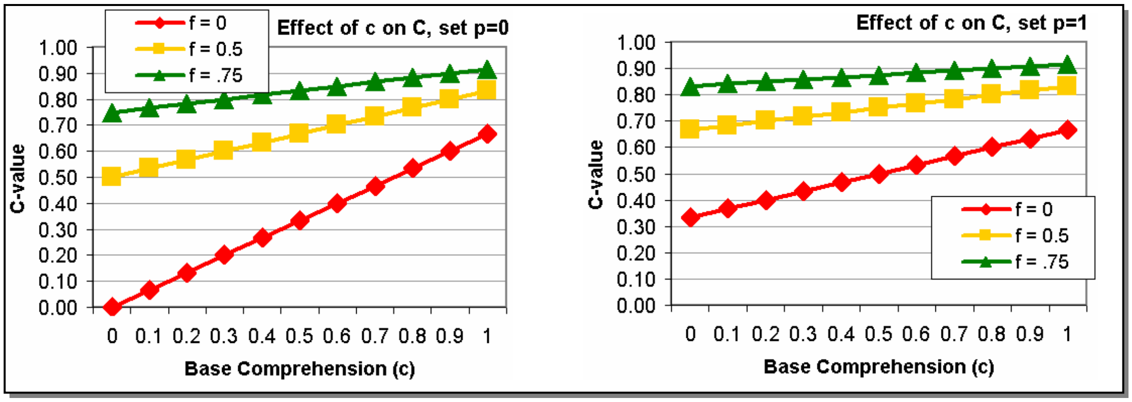
- Still under review is the issue of whether COI should be an additional factor in the C equation. Some justification for this is present in Endsley's argument that temporal dynamics play an important part in assessing the comprehension and projection levels. Specifically, she mentions that part of projection requires an understanding of the rate at which information is changing.

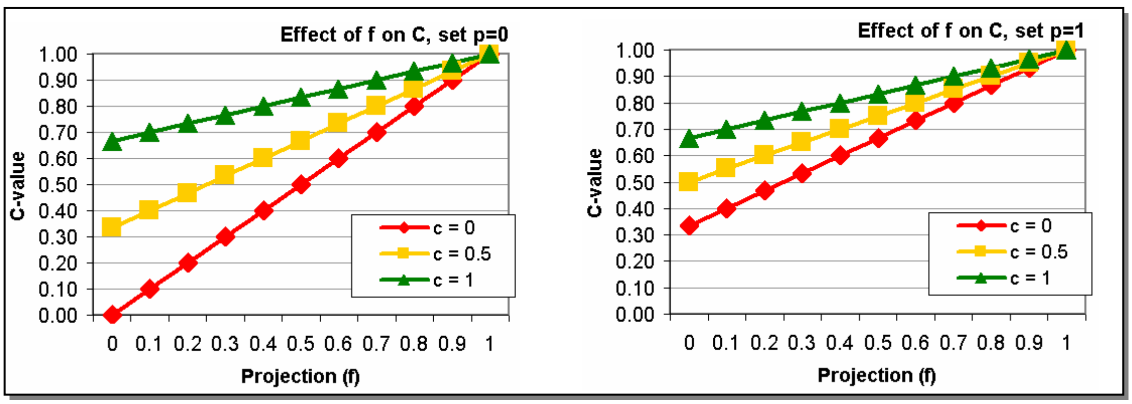
Sample results:

f	c	p	**C**	Rationale
0	0	0	0	
0	0	0.5	0.17	
0	0	1	0.33	Full perception, with no interpretation—no long term awareness or relating notification content to user goals
0	.5	0	0.33	
0	.5	0.5	0.42	
0	.5	1	0.50	
0	1	d*	0.67	Full comprehension of single elements, but no ability to connect elements together to project a trend
.5	.5	0	0.67	An unlikely case that should be prohibited by experimental protocol
.5	.5	0.5	0.71	
.5	.5	1	0.75	
.5	1	d	0.83	c=0 and/or p=0 should be prohibited by experimental protocol
1	d	d	1	

* d = "don't care": any value produces same result




Effect of c on C, set p=0
Effect of c on C, set p=1
f = 0
f = 0.5
f = .75
C-value
Base Comprehension (c)


Effect of f on C, set p=0
Effect of f on C, set p=1
c = 0
c = 0.5
c = 1
C-value
Projection (f)

Acknowledgements

Components of the LINK-UP vision were also developed by John Booker and the students in Spr 2003 CS 6724. Modeling of the IRC equations was a group effort that included John Booker, Maxim Moldenhauer, Ali Ndiwalana, and Edwin Bachetti.

References

1. M. J. Adams, Y. K. Tenney, and R. W. Pew. "Situation Awareness and the Cognitive Management of Complex Systems," *Human Factors 37*, pp. 87-104, 1995.

2. Mica R. Endsley, Betty Bolte, and Debra G. Jones. *Designing for Situation Awareness: An Approach to User-Centered Design*. Taylor & Francis, New York, 2003.

3. D. M. Green and J. A. Swets. *Signal Detection Theory and Psychophysics.* Wiley, New York, 1966.

4. E. Horvitz and J. Apacible. "Learning and Reasoning about Interruption." In *Proceedings of the Fifth International Conference on Multimodal Interfaces*, November 2003, Vancouver, BC, Canada.

5. D. Scott McCrickard, C. M. Chewar, Jacob P. Somervell, and Ali Ndiwalana. "A Model for Notification Systems Evaluation—Assessing User Goals for Multitasking Activity." To appear in *ACM Transactions on Computer-Human Interaction (TOCHI)*, December, 2003. Available at: http://csgrad.cs.vt.edu/~cchewar/paper20.pdf

6. D. Scott McCrickard, Richard Catrambone, C. M. Chewar, and John T. Stasko. "Establishing Tradeoffs that Leverage Attention for Utility: Empirically Evaluating Information Display in Notification Systems." *International Journal of Human-Computer Studies*, volume 8, issue 5, May 2003, pages 547-582.

7. Donald A. Norman. "Cognitive engineering." In Donald A. Norman and Stephen W. Draper, editors, *User Centered System Design: New Perspectives on Human Computer Interaction*, pp. 31–62. Lawrence Erlbaum Associates, 1986.

8. Mary Beth Rosson and John M. Carroll. *Usability Engineering: Scenario-Based Development of Human-Computer Interaction*. Morgan Kaufman, New York, NY, 2002.

9. Alistair Sutcliffe. *The Domain Theory: Patterns for Knowledge and Software Reuse*. Lawrence Erlbaum Associates, Inc., Mahwah, NJ, 2002.

10. David Tessendorf, C. M. Chewar, Ali Ndiwalana, Jon Pryor, D. Scott McCrickard, and Chris North. "An Ordering of Secondary Task Display Attributes." In *Conference Extended Abstracts on Human Factors in Computing Systems (CHI '02)*, pp. 600–601. ACM Press, 2002.

11. Christopher D. Wickens and Justin G. Hollands. *Engineering Psychology and Human Performance*. Prentice Hall, Upper Saddle River, NJ, third edition, 2000.