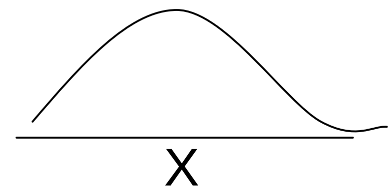
Visualizing Uncertainty in High Time-Stress

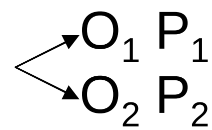
Christopher D. Wickens
Head, Human Factors Division
University of Illinois, Institute of Aviation

- Terms
- Model of Influences
- Challenge and Results of Empirical Research
- Human Factors Guidelines for Best Display Practices

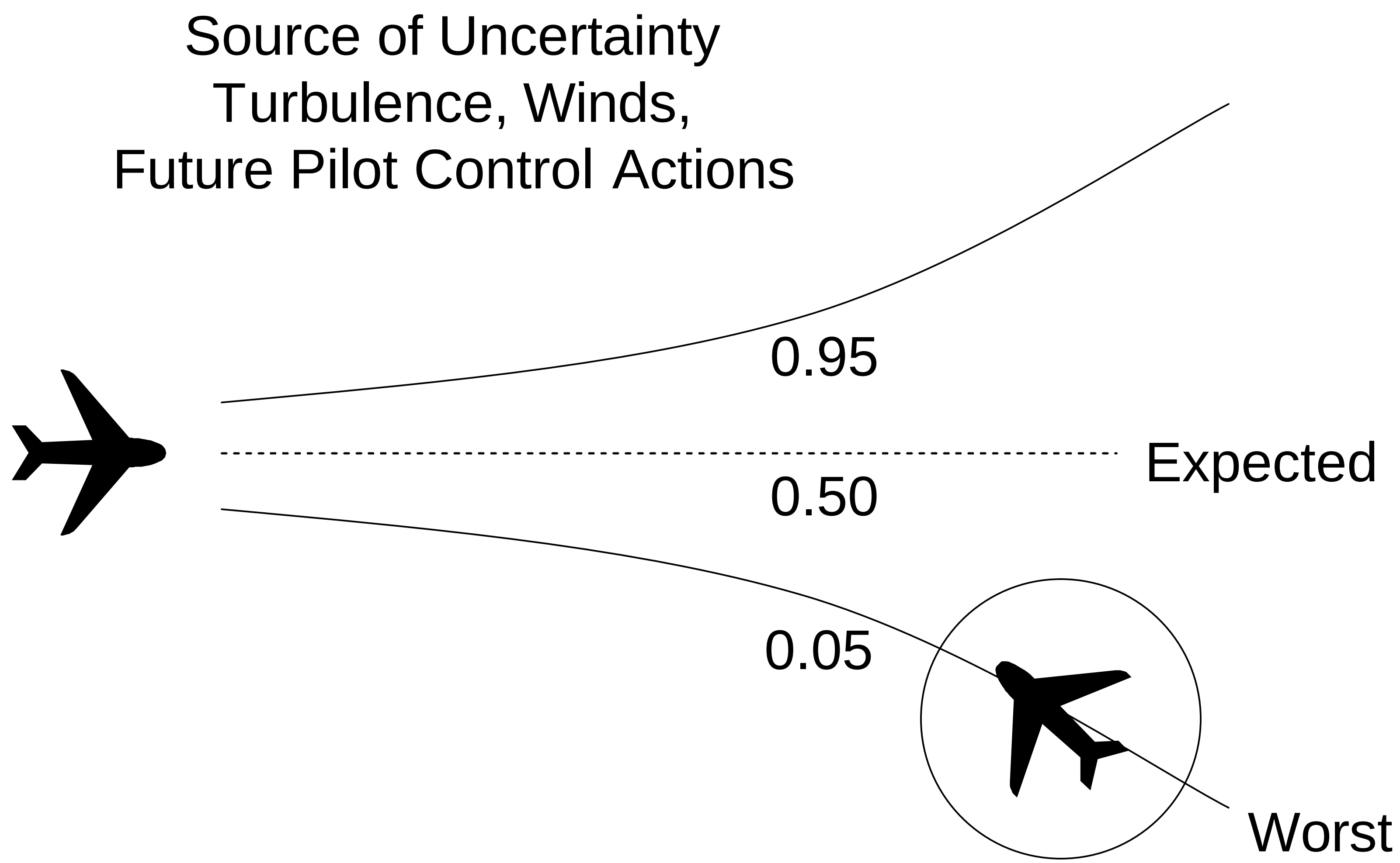
Visualizing Uncertainty in High Time-Stress

- Uncertainty:
  Spatial-temporal resolution (imprecision)
  categorical uncertainty

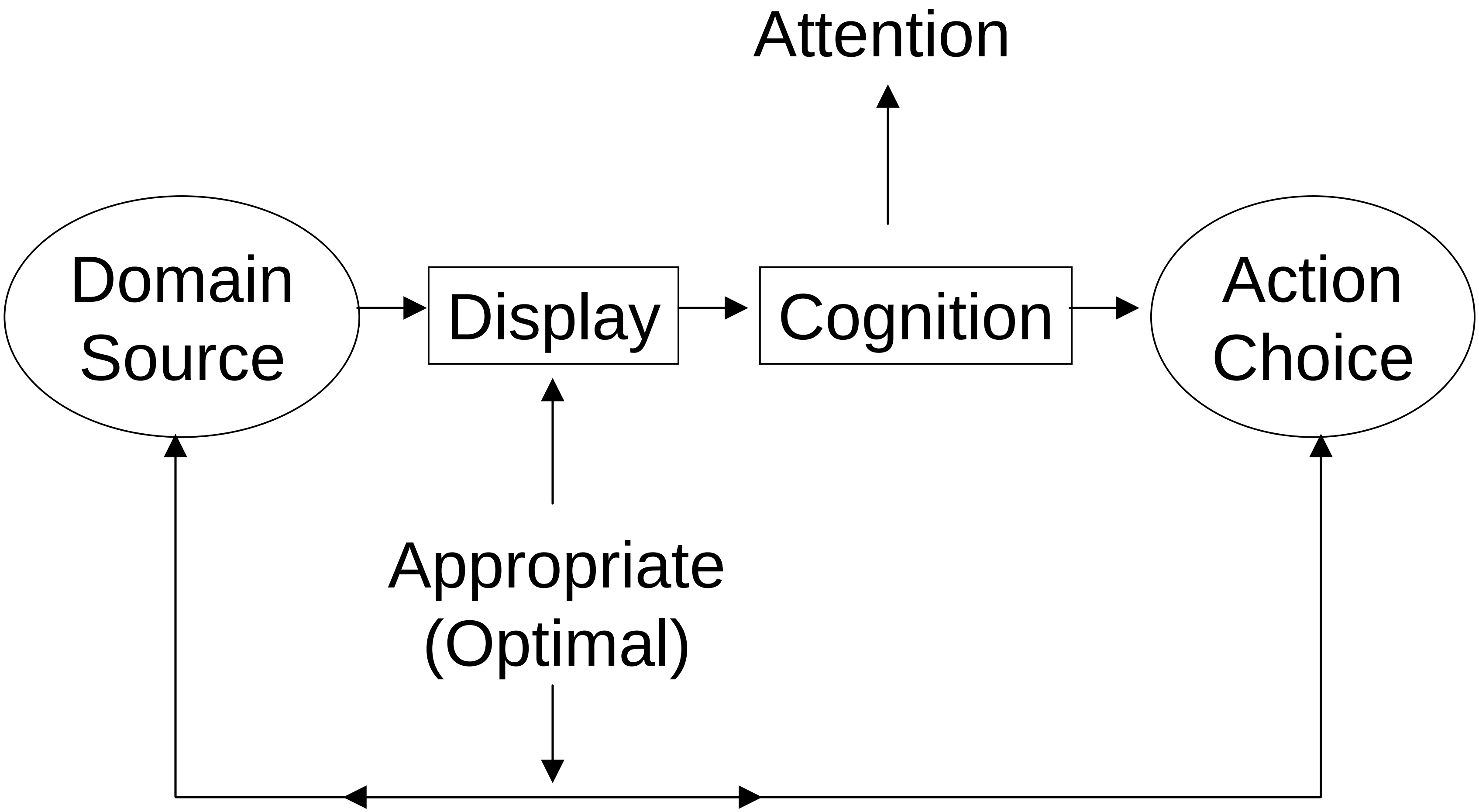


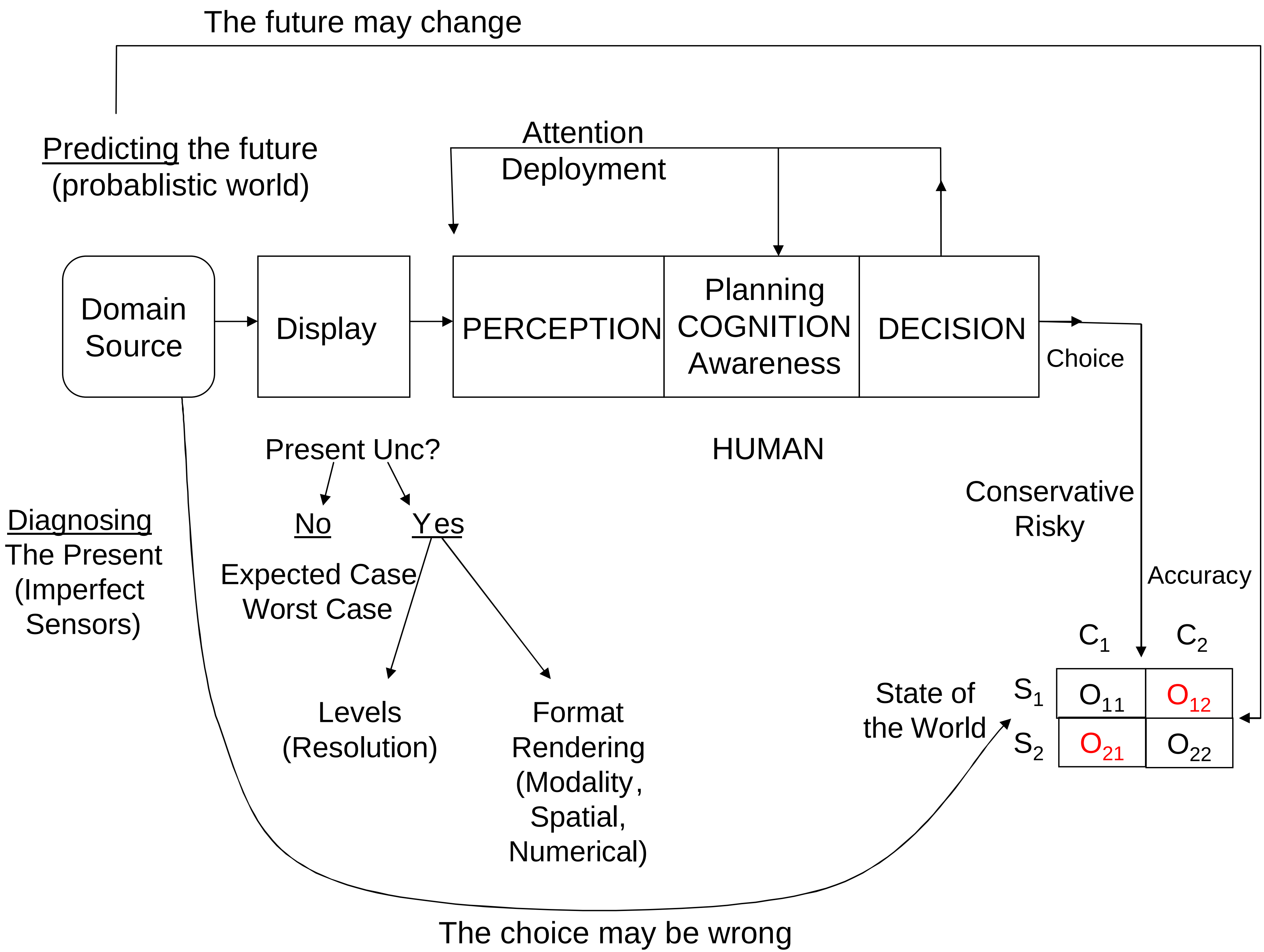


- Expected Value (Risk) PX**V**
  People relatively poor at utilizing probablistic information. Value dominates risk decisions.
- Visualizing: Displays. Multi-media (visual, auditory sounds). Not linguistic.
- Time Stress: Minimizing cognitive load: best outcome
- Example: the pilot conflict avoidance maneuver.


Source of Uncertainty
Turbulence, Winds,
Future Pilot Control Actions
0.95
Expected
0.50
0.05
Worst

A Simple Model




The future may change
Predicting the future
(probablistic world)
Attention
Deployment
Domain
Source
Display
PERCEPTION
Planning
COGNITION
Awareness
DECISION
Choice
Present Unc?
No
Yes
HUMAN
Conservative
Risky
Diagnosing
The Present
(Imperfect
Sensors)
Expected Case
Worst Case
Accuracy
C_1
C_2
State of
the World
S_1
S_2
O_{11}
O_{12}
O_{21}
O_{22}
Levels
(Resolution)
Format
Rendering
(Modality,
Spatial,
Numerical)
The choice may be wrong

Empirical Research on “What Works”: The Challenge

Make credible the actuarial experience of probabilities.

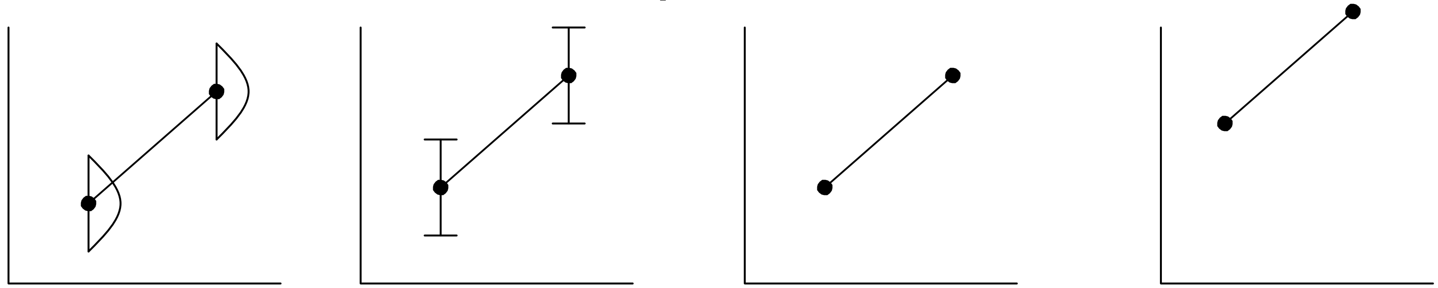
If low probability events are part of the display rendering, they must be experienced by the participant.

Rendering of p=.01 event, participant must (a) experience the event, (b) experience it 1 out of 100, or (better) 2 out of 200.

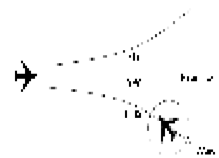
Few studies exist that have:

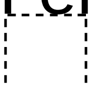
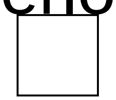
* compared uncertainty representation vs. none.
* compared different formats of uncertainty representation
* collected <u>objective</u> performance data with actuarial experience

The Empirical Results



- Display: Uncertainty vs. “expected case” or “worst case”
- No effect? Wickens Gempler & Morphew. Probablistic display of predicted flight path error does not help conflict avoidance.



- Yeh, Merlo & Wickens. Uncertain intelligence template  vs.  does not improve attention allocation in military target cueing when  explicitly displays degraded spatial resolution (increased position uncertainty) of the cue.
- Smith & Wickens: Highlighting best case, expected case, worst case outcomes does not alter NMD strategic missile launch decisions

Empirical Results: Best Display Practices

1. Levels of resolution: (Danger, Uncertain, Safe)

(2) D-S (3) D-U-C (5) D- DU U US S

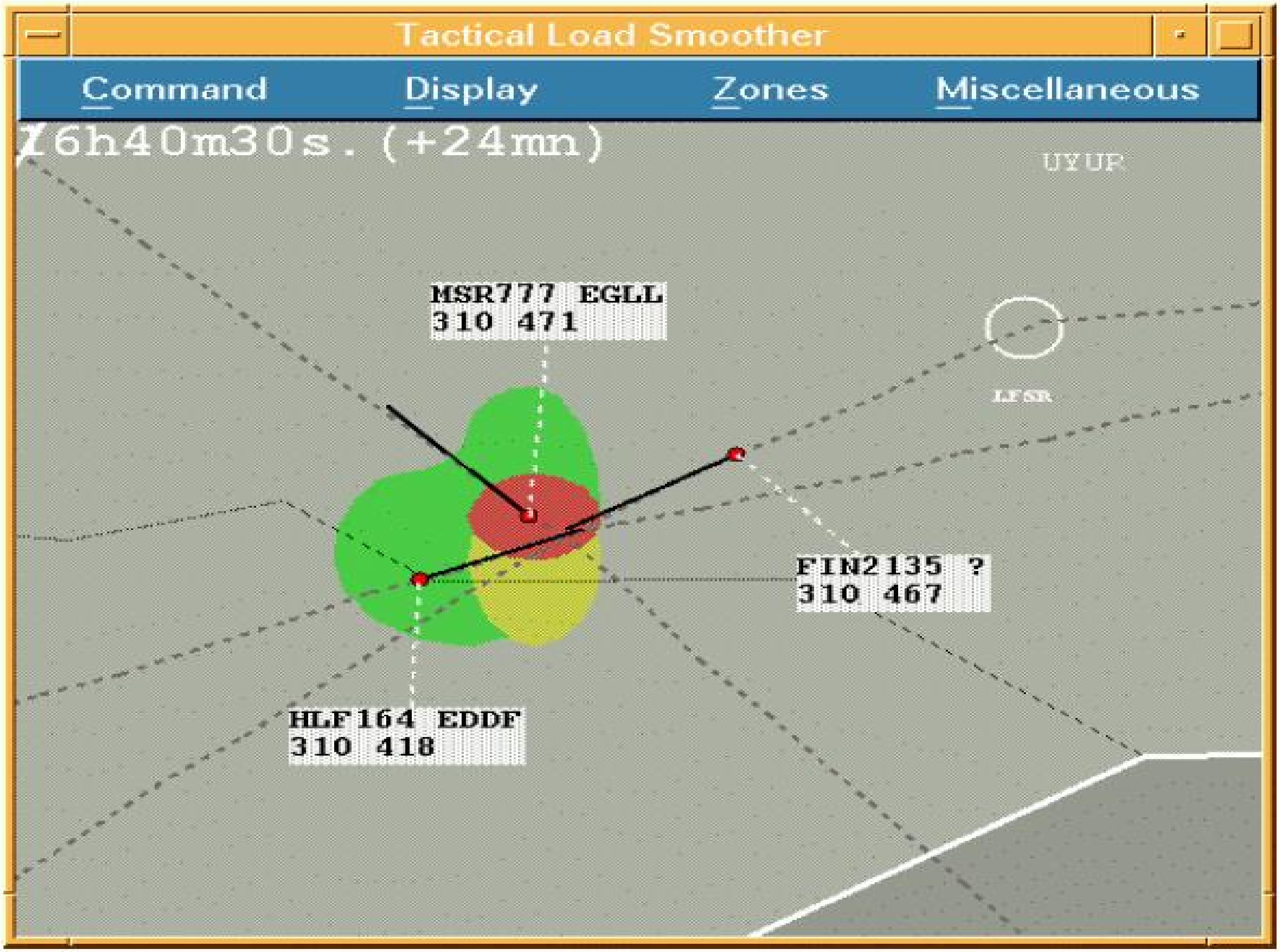
More (than 2) levels help. (St. Johns and Mannes, Schinzer et al). Philosophy of "likelihood alarm". (Sorkin & Woods). Aviation collision warnings. Why? More of the errors in a higher resolution system are not as "bad". (This fosters greater trust in the system):

	D	S
D		**L**
S	**L**	

	D	U	S
D		**L**	
U	**L**		**L**
S		**L**	

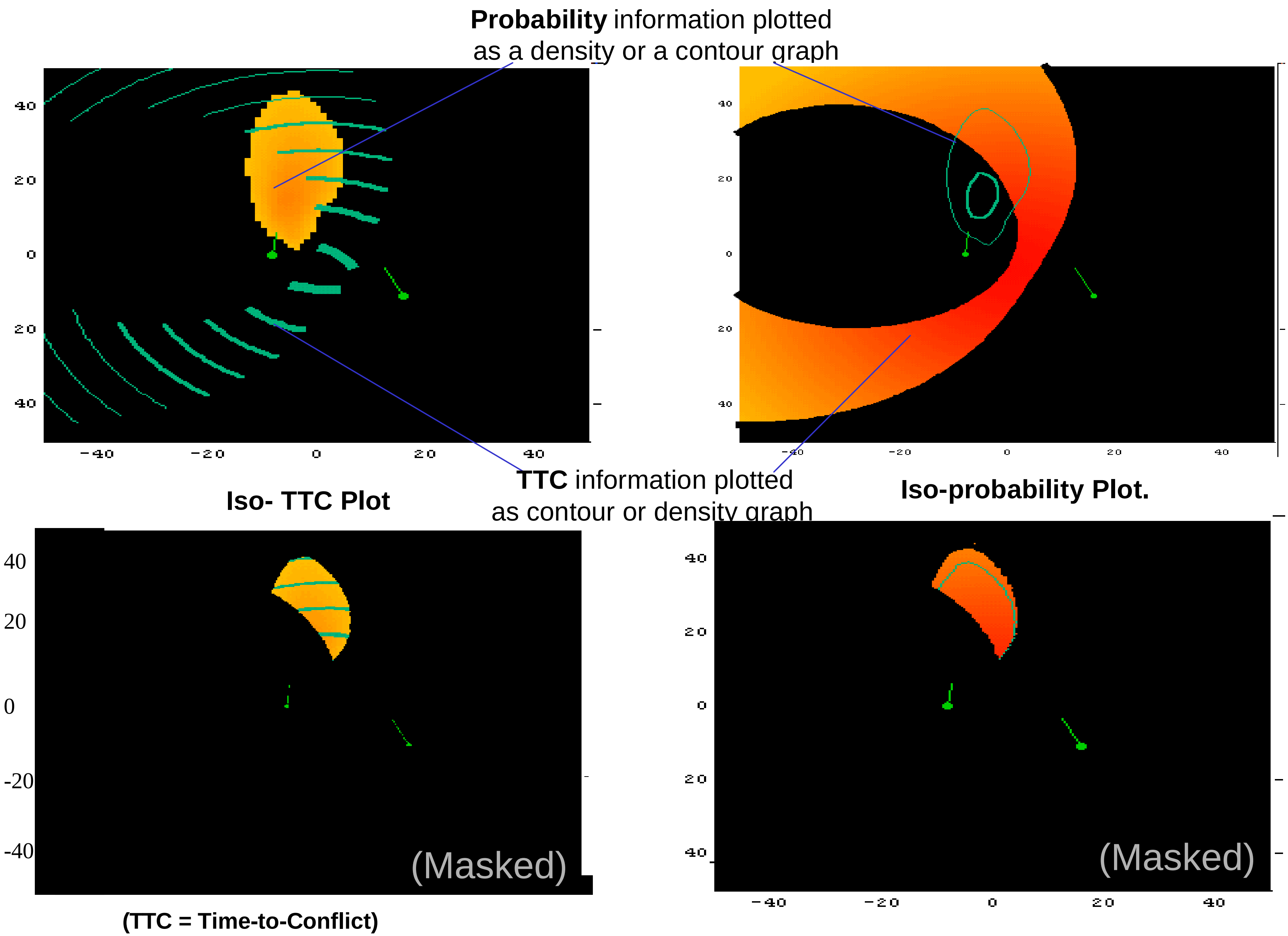
How many levels needed? (Schinzer). > 4 may be all.

How to render?


Tactical Load Smoother
Command
Display
Zones
Miscellaneous
16h40m30s. (+24mn)
UYUR
MSR777 EGLL
310 471
LFSR
FIN2135 ?
310 467
HLF164 EDDF
310 418

Predictive Probabilistic and Temporal Conflict Avoidance Displays

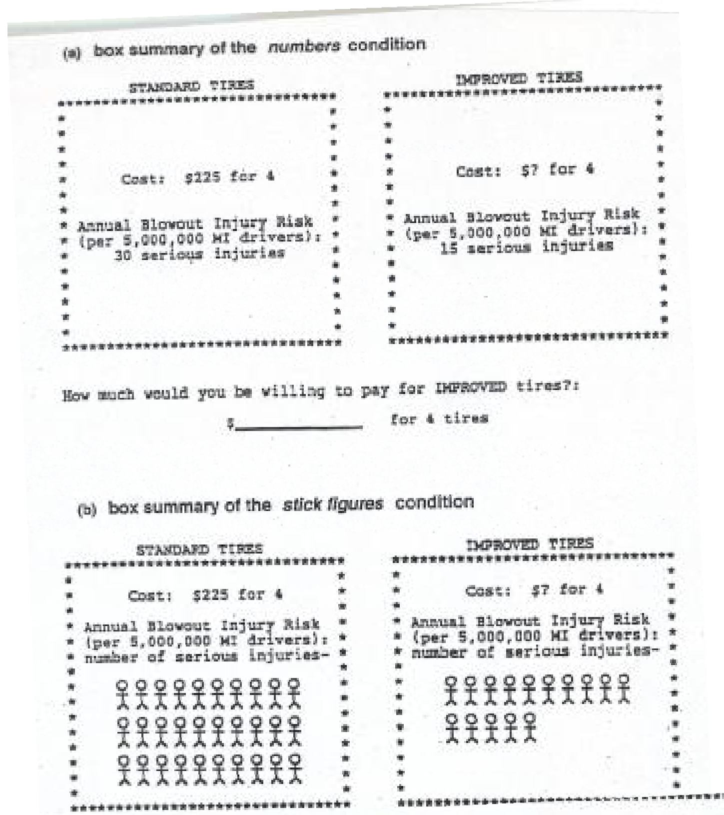
(courtesy of Jason Telner & Paul Milgram, University of Toronto)



Empirical Format Comparisons

- Graphical vs. Numerical/verbal. Graphical wins **J**
  (Stone et al, Kirschenbaum & Aruda, Andre & Cutler, Kirlik & Nunes)
  but not always (Schinzer et al)
- Visual vs. Auditory, Tactile: Visual wins **J** (Basapur)
- Visual Spatial vs. Visual color: Spatial wins **J** (Andre). But not always (Schinzer et al).

Stone et al.



(a) box summary of the *numbers* condition

STANDARD TIRES	IMPROVED TIRES
Cost: $225 for 4	Cost: $7 for 4
Annual Blowout Injury Risk (per 5,000,000 MI drivers): 30 serious injuries	Annual Blowout Injury Risk (per 5,000,000 MI drivers): 15 serious injuries

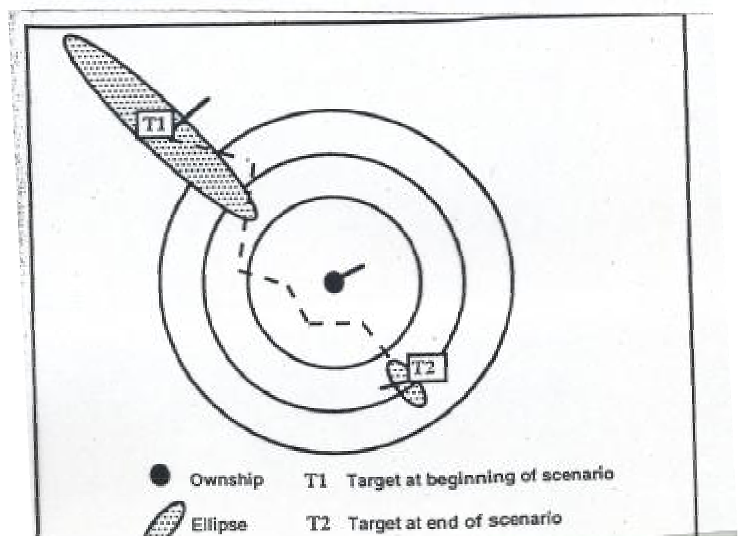
How much would you be willing to pay for IMPROVED tires?:

$__________ for 4 tires

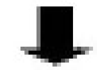
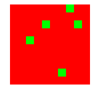
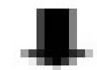
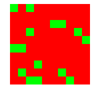
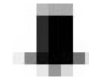
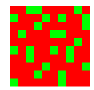
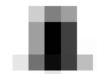
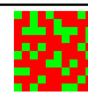
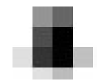
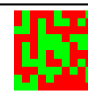
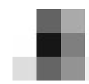
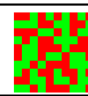
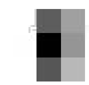
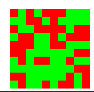
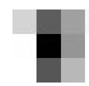
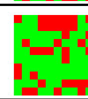
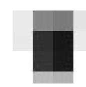
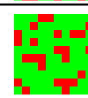
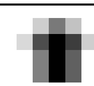
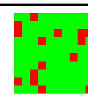
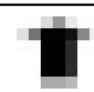
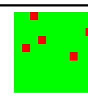
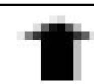
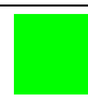
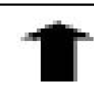
(b) box summary of the *stick figures* condition

STANDARD TIRES	IMPROVED TIRES
Cost: $225 for 4	Cost: $7 for 4
Annual Blowout Injury Risk (per 5,000,000 MI drivers): number of serious injuries-	Annual Blowout Injury Risk (per 5,000,000 MI drivers): number of serious injuries-

Kirchenbaum & Aruda



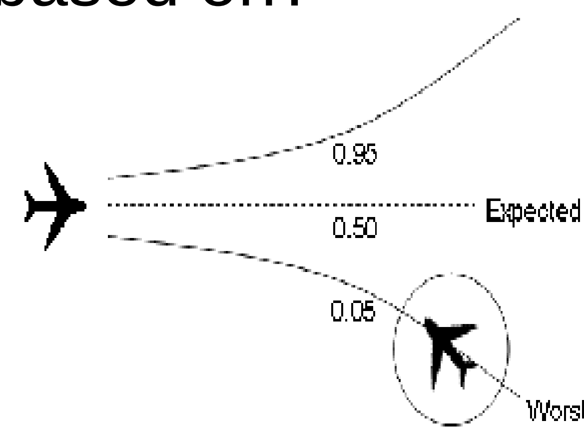
Schinzer et al: Investment Decisions

Range (High)	Numeric Expression	Linguistic Expression	Colored Icon	Arrow Icon	
0	0%	Absolutely Impossible			***
0-.9	5%	Rarely			
.9-.18	14%	Very Unlikely			**
.18-.27	23%	Fairly Unlikely			*
.27-.36	32%	Somewhat Unlikely			**
.36-.45	41%	Uncertain			
.45-.54	50%	Tossup			***
.54-.63	59%	Better Than Even			
.63-.72	68%	Rather Likely			**
.72-.81	77%	Quite Likely			*
.81-.90	86%	Highly Probable			**
.90-1	95%	Almost Certain			
1.0	100%	Absolutely Certain			***

Best Practices in Time Stressed Environments

- Cognitive limitations: (Sweller: Cognitive load theory) Limited time, limited expertise
- Extensive research on graphical presentation (Tufte, Gillen et al., Wickens & Hollands)
- Information overload: people will filter: what will they process? Ignore?
- What will decision be based on?

  Expected case? Worst case?
- What <u>should</u> decisions be based on?

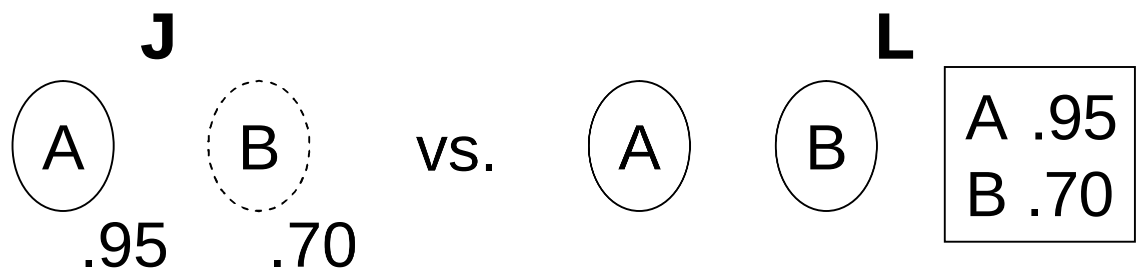


Best Practices Under Time Stress

1. Eliminate redundant extra information (declutter)

Just as good?

2. Visually link uncertainty representation to uncertain element (Proximity compatibility principle): Why visual display is good.



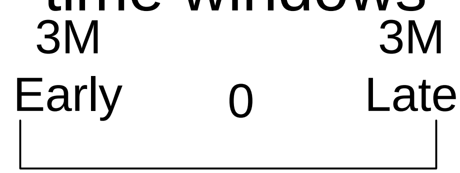
3. Express uncertainty in the "language of action" for:

DIAGNOSIS

Spatial occupancy contours

PREDICTION

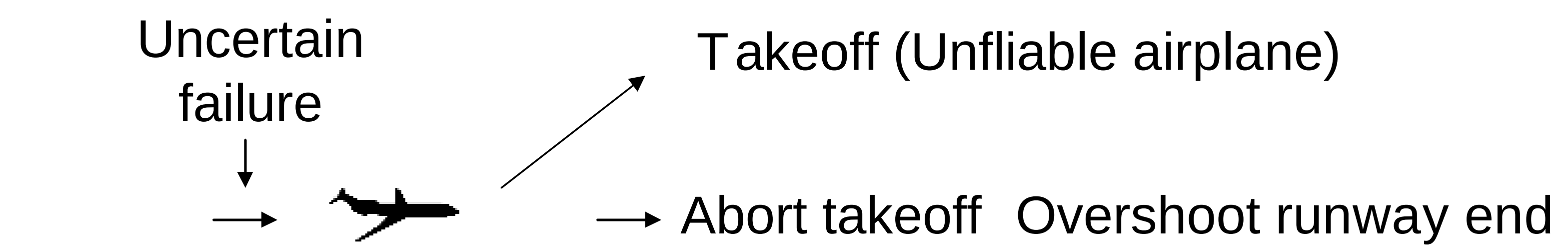
time windows



4. Need for standardization of contour level (95%?)

Consequences of Supporting Risk-Seeking vs. Risk Aversive Behavior

- What kind of behavior does displaying uncertainty induce, invite? 1. That uncertainty exists. 2. How big it is.
- In high time pressure designer should evaluate the worst case outcomes. Design to avoid these, presenting relatively less probablistic information as time pressure grows.
- The aborted takeoff decision in aviation (Inagake).



Conclusions

More research needed (Echoes calls by others)

Analyze consequences of human knowing uncertainty

Displaying Information will induce specific behavior in high time-stress

References

Andre, A. D., & Cutler, H. A. (1998). Displaying uncertainty in advanced navigation systems. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting.* Santa Monica, CA: Human Factors Society.

Basapur, S., Bisantz, A. M., & Kesavadas, T. (2004). The effect of display modality on decision-making with uncertainty. *Proceedings of the Human Factors and Ergonomics Society. Annual Meeting.* Santa Monica, CA: Human Factors Society.

Bisantz, A. M., Schinzing, S., & Munch, J. Uni displaying uncertainty: Investigating the effects of display format and specificity. Unpublished MS.

Finger, R. & Bisantz, A. M. (2002). Utilizing graphical formats to convey uncertainty in a decision making task. *Theoretical Issues in Ergonomics Science, 3*(1), 1-25.

Gillan, D.J., Wickens, C.D., Hollands, J.G., & Carswell, C.M. (1998). Guidelines for presenting quantitative data in HFES publications. *Human Factors, 40(1), 28-41.*

Kirschenbaum, S. S., & Arruda, J. E. (1994). Effects of graphic and verbal probability information on command decision making. *Human Factors, 36*(3), 406-418.

Lipkus, I. M., & Hollands, J. (2000). The visual communication of risk. Monographs of the National Cancer Institute. Bethesda, MD: National Institutes of Health.

Nicholls, D., Battino, P., Marti, P., & Pozzi, S. Presenting uncertainty to controllers and pilots.

Sarter, N. B., & Schroeder, B. (2001). Supporting decision-making and action selection under time pressure and uncertainty: The case of inflight icing. *Human Factors, 43*, 580-590.

Smith, M., & Wickens, C. D. (1999). *The effects of display highlighting and event history on operator decision making in a National Missile Defense system application* (ARL-99-7/FED-LAB-99-4). Savoy, IL: University of Illinois, Aviation Research Laboratory.

Wickens, C. D. (1996). Designing for stress. In J. Driskell & E. Salas (Eds.), *Stress and human performance* (pp. 279-295). Mahwah, NJ: Lawrence Erlbaum.

Wickens, C. D., & Carswell, C. M. (1995). The proximity compatibility principle: Its psychological foundation and its relevance to display design. *Human Factors, 37*(3), 473-494.

Wickens, C. D., & Dixon, S. (2005). *Is there a magic number 7 (to the minus 1)? The benefits of imperfect diagnostic automation: A synthesis of the literature* (AHFD-05-1/MAAD-05-1). Savoy, IL: University of Illinois, Aviation Human Factors Division.

Wickens, C. D., Gempler, K., & Morphew, M. E. (2000). Workload and reliability of predictor displays in aircraft traffic avoidance. *Transportation Human Factors Journal, 2*(2), 99-126.

Wickens, C. D., & Hollands, J. (2000). *Engineering psychology and human performance* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Yeh, M., Merlo J. L., Wickens, C. D. & Brandenburg, D. L. (2003). Head up versus head down: The costs of imprecision, unreliability, and visual clutter on cue effectiveness for display signaling. *Human Factors, 45(3), 390-407.*