

Wizard of Oz (WoZ)—A Yellow Brick Journey

John F. ("Jeff") Kelley
Master Inventor / Sr.
Managing Consultant
IBM iX
130 Laynewade Rd
Polk City, FL
USA
jfkcxp@hfergo.com

We're off to see The Wizard. [Dorothy Gale and her posse]

Since 1983, the term "Wizard of Oz" has come into common usage in the fields of Experimental Psychology, Human Factors, Ergonomics and Usability Engineering to describe a testing or iterative design methodology wherein an experimenter (the "Wizard"), in a laboratory setting, simulates the behavior of a theoretical intelligent computer application (often by going into another room and intercepting all communications between participant and system). Sometimes this is done with the participant's a priori knowledge, and sometimes it is a low-level deceit employed to manage the participant's expectations and encourage natural behaviors (though always, I would hope, with appropriate disclosure during the debriefing part of the experiments!).

I coined that term around 1980 to describe the methodology I developed during my dissertation work at Johns Hopkins University. My dissertation advisor was Professor Alphonse Chapanis, the "Godfather of Human Factors and Engineering Psychology." This column is, justifiably, just as much about AI (the *real* Wizard in this story) as it is about WoZ (or me).

In 1970, Chapanis started the Communications Research Laboratory (CRL) at Johns Hopkins University. In the 12 years that followed, he and his graduate students did seminal work in the nascent field of multimodal communication: technology-mediated human-human communications and human-computer communications using video, audio, handwriting, and keyboard. This work led to significant advances in telecommunications, teleconferencing, video conferencing and the intelligibility of digitized speech. Among many significant advances to come out of CRL was this "Wizard of OZ" method for development and assessment of natural language recognition systems that are in wide use today.



There had been previous “experimenter in the loop” studies in Chapanis’ Communications Research Lab, notably Randy Ford’s CHECKBOOK study which examined the differences between spoken and typed natural language. My dissertation explored the use of this experimenter intervention to train the software to recognize natural language inputs on its own.

Pay no attention to the man behind the curtain! [The Wizard, when exposed by Toto]

There are several interesting tidbits in this column that you may not find in the literature...

Tidbit: Amusingly enough, in addition to some one-way mirrors and such, there literally was a curtain separating me, as the “Wizard,” from view by the participant during the study.

Tidbit: I did originally have a definition for the “OZ” acronym (aside from the obvious parallels with the 1900 book by L. Frank Baum). “Offline Zero” was a reference to the fact that an experimenter (the “Wizard”) was interpreting the users’ inputs in real time during the simulation phase and iterating the software toward the goal of a hands-off experience. Folks laughed at this lily-gilding as an expression of my acronymophilia and I eventually dropped it.

From 1959 through 1995, Chapanis had a long-term consulting relationship with IBM. In addition to setting up training programs to educate executives, managers, engineers, and programmers about man-machine interface considerations (human-computer interaction) in the design of computing technologies, he also received support from IBM (and later, from GTE) for various communications graduate student research projects.

Our primary sponsor for my dissertation work was John D. Gould from IBM’s T. J. Watson Research Center. John is known for his contributions to applying human factors to the principles of system design, as well as his seminal work on how authors think about their writing, dictating, and speaking.

Tidbit: John hired me after my dissertation work and I spent 18 happy years at IBM Research before embarking on my subsequent (and also happy) career as an IBM Usability Engineering Consultant.

I’ll get you my pretty, and your little dog too! [The Wicked Witch of the West, after Dorothy’s house landed on her sister]

Tidbit: In my publication (Kelley, 1984), I referred to the fact that, in the early phases of the experiment, “the experimenter simulated the recognition system *in toto*.” The journal editor, Stu Card, only dictated one change to my manuscript: that I not *italicize* the frivolous pun (*in toto*). I said “fine” and the typesetters, as they usually do, ended up italicizing the Latin anyway. Thus, the best pun of my career ended up italicized in a scientific journal!

What would you do with a brain if you had one? [Dorothy Gale to The Scarecrow]

The OZ methodology may not be as powerful as the “great and powerful” wizard, but it is very powerful nonetheless.

In its original application, I was able to create a simple keyboard-input natural language recognition system that far exceeded the recognition rates of any of the far more complex systems of the day. The key enabling factor was that the system was designed to work in a single context (calendar-keeping), which constrained the complexity of language encountered from users to the extent where a simple language processing model was sufficient to meet the goals of the application. I called this an “empirical grammar.” The processing model was a two-pass keyword/keyphrase matching approach, based loosely on the algorithms employed in Weizenbaum’s famous parody of the computerized psychotherapist: Eliza (Weizenbaum, 1976).

OZ was important because it addressed the obvious criticism: Who can afford to use an iterative methodology to build a separate natural language system (dictionaries, syntax) for each new context?

The answer turned out to be—by using an empirical approach like OZ, anyone can afford to do this. My dictionary and syntax growth reached asymptote (better than 90% recognition) after only 16 experimental trials, and my resulting program, with dictionaries, was less than 300k of (APL) code.

My! People come and go so quickly here! [Dorothy to herself]

I'll confess that my nurturing of WoZ after publishing my dissertation (Kelley, 1983; Kelley 1984) can only be charitably described as "benign neglect." My career has tended more toward UCD and high-fidelity, rapid prototyping (Kelley, 2007; Kelley, 2012).

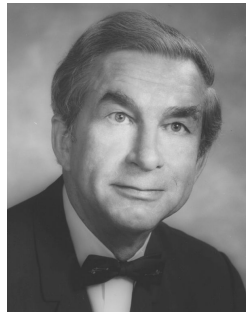
I was recently pleasantly surprised to find 125,000 Google results for "wizard of oz" usability, including many retrospective articles. ACM shows 301 research papers with "Wizard of Oz" (52 with that in the title). My own two publications have 150 citations and 3,051 downloads (including 500 in the last year). And people do come up to me at conferences to chat about it.

Aside from the obvious use of WoZ in straight-up usability testing for software that isn't quite ready for prime time, it has found a lot of use in training NLP/IVR systems (collecting representative samples with a phone system that only appears to automatically recognize spoken text), and even in my own company's IBM Design Thinking methodology and Watson AI systems (who knew?). One of my favorite anecdotes was the pizza delivery truck that was supposedly fully automated, but had a human driver hidden in a box covering the passenger seat! Paul Green at University of Michigan has pioneered a lot of this work. There is also interesting work coming out of Stanford (see Wang, Sibi, Mok, & Ju, 2017).

There's no place like home. [Dorothy to herself]

The Tin Man might just as well have said: "Home is where the heart is," and we practitioners of UCD chose this field because that's where our hearts led us.

I owe my own calling to my mentor, the late, great Al Chapanis. Here is some of what I wrote for his eulogy:



Alphonse Chapanis was born in Meriden, CT USA on 17 March 1917 and passed away on 4 October 2002 in Baltimore, MD USA. Dr. Chapanis received his Ph.D. in Psychology from Yale University in 1943. Aside from 22 years of consultancy in Human Factors and Ergonomics, Dr. Chapanis served as a 2nd Lieutenant at the U.S. Army's Aero Medical Laboratory and was Professor of Experimental Psychology at The Johns Hopkins University from 1946 through 1982.

Alphonse Chapanis has been warmly described as the "Godfather of Human Factors and Ergonomics." His contributions to the fields of applied experimental psychology, human factors, and ergonomics spanned a career of over 60 years.

The 95 books and articles cited in his memoir [Chapanis, 1999] are only a portion of publications he has authored, many of them significant in the evolution of the fields of human factors and ergonomics. Much of his work displays his interest in merging the fields of basic psychological research (notably vision and perception) with the practical application of research to

engineering design and led to his being recognized as a distinguished leader in the fields of human factors engineering and ergonomics.

Chapanis [and colleagues] wrote the first human factors ergonomics textbook [Chapanis, Garner, & Morgan, 1949] and wrote an award-winning Scientific American article on color blindness [Chapanis, 1951]. These were partially a result of some ground-breaking work he did for the U.S. Army's Aero Medical Laboratory [AML] during World War II, as well as subsequent systems research conducted at a U.S. Navy sponsored field laboratory at The Johns Hopkins University.

While completing his Ph.D. in Psychology at Yale, Chapanis joined AML as a 2nd Lieutenant and began training in aviation psychology. His early work involved the design of night-vision displays and the effects of anoxia and high g-forces on vision loss.

In one famous study, Chapanis was studying the high incidence of post-landing crashes of the B-17 "Flying Fortress" aircraft. Not satisfied with the common classification of "pilot error," Chapanis studied the incidents and concluded that the problem was, instead, a matter of "designer error." As the controls for the flaps and the landing gear were identical and placed side-by-side, tired pilots landing at night after a long flight could easily select the wrong control. Chapanis solved the problem by "shape-coding" the controls (one resembling a wheel, the other resembling a flap), and providing a fail-safe control connected to pressure sensors in the landing gear (preventing retraction if the aircraft's weight was on the gear).

*From 1946 until 1982, Professor Chapanis was a prolific member of the Psychology Department faculty at The Johns Hopkins University, graduating, among others, 29 Ph.D.'s. In the early years, Chapanis, working under contracts from the U.S. Navy, made fundamental contributions to the emerging field of systems engineering. His book *Research Techniques in Human Engineering* [Chapanis, 1965a] had a great impact on the field.*

It was also during this time that Chapanis adapted a critical incident approach to safety issues (such as hospital medication errors) and wrote a widely-cited paper entitled "Words, Words, Words" [Chapanis, 1965b] in which he decried the prevalence of confusing and conflicting wording in signs and labels in contexts ranging from highways to instruction placards to elevators, often leading to errors and safety issues.

In 1953 and 1954, Chapanis consulted with Bell Labs, doing ergonomics design studies of new toll-operators' keypads that eventually contributed to the design of the push-button telephone.

In the 1960s, Chapanis was an active international proponent of human factors and ergonomics which included lectures, and publications about cross-cultural issues in human factors and ergonomics. He was elected to the Council of the International Ergonomics Association [IEA] in 1967 and became president of that organization in 1973. There are interesting anecdotes in his memoir concerning his role in reporting observations from his travels, especially behind the Iron Curtain, back to the Office of Naval Research and the U.S. military.

Alphonse Chapanis was actively involved in professional organizations throughout his career. In addition to his role with the IEA, Alphonse Chapanis was the 1960 President of the Society of Engineering Psychologists [Division 21 of the American Psychological Association] and the 1964 President of the Human Factors and Ergonomics Society [HFES]. In 1982 he won the IEA's Distinguished Service Award and, in 1987, the HFES President's Distinguished Service Award. He was also an advocate for professional qualification standards, helping to establish the Board of Certification in Professional Ergonomics in the mid-1980s.

Professor Chapanis, always a fierce advocate of his students, is remembered with great fondness by the many graduate students he shepherded into successful careers, despite (or, perhaps, because of) his legendary attention to detail and insistence on clear writing style. In recognition of his contributions in education, the Human Factors and Ergonomics Society, in 1983, renamed its prestigious Best Student Paper Award the "Alphonse Chapanis Award."

Some people without brains do an awful lot of talking, don't they? [The Scarecrow to Dorothy]

Tidbit: As evidenced in his article "Words, Words, Words" (Chapanis, 1965b), Al was a real stickler for not using 25-cent words when simple language would do. ("Don't say 'utilize', say 'use!'") My joke in grad school was that we got our marked-up drafts back from The Professor weighing 20% more from all the pencil graphite he would add.

Tidbit: Well, I got my own back on the great man: In my personal dissertation acknowledgement page (added after my defense of the final draft), I wrote "If this paper *eschews sesquipedalian obfuscation*, that is only because of Professor Chapanis' patient and persistent efforts over the last four years to teach me to think and write clearly." Al was so tickled by this, he included it in his memoirs.

We're not in Kansas anymore. [Dorothy to Toto]

Tidbit: Did Alphonse Chapanis perform espionage against Stalin's USSR for a U.S. intelligence agency? (Go read his memoirs and draw your own conclusions!)

I think I'll miss you most of all! [Dorothy to the Tin Man]

In my HFES Presidential Address (Kelley, 2008), I quoted Professor Chapanis who said in his memoirs (Chapanis, 1999): "*There is one thing I have never regretted - and that is my choice of profession. Human factors has always been challenging, frustrating at times, rewarding at others, but never dull. I can honestly say in retrospect that I have had a full life - an exciting life - and that I have enjoyed telling people about human factors, educating students and others to take over where I have had to leave off, and grappling with problems of trying to make our material world safer, more comfortable, and easier to cope with. In fact, there is only one thing I truly regret - I'm sorry I've come to the end*" (p. 234).

I fervently hope that I am not, yet, at the end of **my** career, but when the time comes, I'll be able to retire a happy man if I had earned the right to say that by applying what I know, I've left some small corner of the world a teensy bit more of a congenial place to live in.

Thank you for reading. I'd love to hear your thoughts and stories about WoZ or Al Chapanis: jfkcuxp@hfergo.com

References

- Chapanis, A. (1951). Color blindness. *Scientific American*, 184, 48–53.
- Chapanis, A. (1965a). *Research techniques in human engineering*. Baltimore, MD: The Johns Hopkins University Press.
- Chapanis, A. (1965b). Words, words, words. *Human Factors*, 7, 1–17.
- Chapanis, A. (1999). *The Chapanis chronicles: 50 years of human factors research, education, and design*. Santa Barbara, CA: Aegean Publishing Company.
- Chapanis, A., Garner, W. R., & Morgan, C. T. (1949). *Applied experimental psychology: Human factors in engineering design*. New York, NY: Wiley.
- Kelley, J. F. (1983, December 13–15). An empirical methodology for writing user-friendly natural language computer applications. *Proceedings of ACM SIG-CHI '83 Human Factors in Computing Systems* (Boston; pp. 193–196), New York, NY: ACM.

- Kelley, J. F. (1984, March). An iterative design methodology for user-friendly natural language office information applications. *ACM Transactions on Office Information Systems*, 2(1), 26–41.
- Kelley, J. F. (2007, November 8). Agile + usability engineering: Strange bedfellows? Keynote Address: World Usability Day 2007 Conference, Dayton, OH.
- Kelley, J. F. (2008). *HFES presidential address*, Human Factors and Ergonomics Society Annual Meeting, New York, NY. Available from <http://jfkprez.blogspot.com>
- Kelley, J. F. (2012). Over the waterfall in a barrel—Adapting agile techniques to user experience in a non-agile world. Part of a panel: User Experience and Agile Development. *Proceedings of the Annual Meeting of the Human Factors and Ergonomics Society*, Santa Monica, CA. Available from <http://bit.ly/2kew4RG>
- Wang, P, Sibi, S., Mok, B., & Ju, W. (2017). Marionette: Enabling on-road Wizard-of-Oz autonomous driving studies. *ACM/IEEE International Conference* (pp. 234–243). Available from <http://bit.ly/2H3ERQ8>
- Weizenbaum, J. (1976). *Computer power and human reason: From judgment to calculation* (pp. 2, 3, 6, 182, 189). New York: W.H. Freeman and Company.

About the Author



John F. ("Jeff") Kelley, PhD, CUXP

Dr. Kelley is an IBM Master Inventor (42 patents and 24 invention publications) and certified Sr. Managing Consultant in the Research Insights practice of IBM Interactive Experience. As a Sr. UX architect and team leader, he has been involved in all aspects of client engagements: proposals, visioning, requirements, solution design/delivery, prototyping, UI design innovation, development, and stakeholder/customer interaction.

He pioneered and published user centered design methodologies (for example, the "Wizard of OZ" simulation technique widely used in the HCI community). He also created AgileRemote: an award-winning, user-centered design methodology applying online surveys and Google Spreadsheets to gather/rank Agile User Stories that directly populate conceptual models and design prototypes which export navigation and content directly to development environments in various formats.

He has earned the distinction of holding the following honors: Certified User eXperience Professional (CUXP), Fellow of the Human Factors & Ergonomics Society, and Chartered Fellow of the Institute of Ergonomics & Human Factors.