

Identifying Experts for the Design of Human-Centered Systems

Robert R. Hoffman
IHMC

Abstract

The identification of experts is crucial in many research projects and application areas for intelligent systems, including the development of rational algorithms and the creation of knowledge bases. This essay addresses the question of how to identify experts, offering a method that is more robust and scientifically grounded than the common reliance on the so-called "ten-year" or "10,000 hours" rules for deciding who is, and who is not an expert.

INTRODUCTION

The question of how many experts are sufficient for knowledge elicitation in support of the development of intelligent systems presupposes that one has a robust method for determining who is an expert in the first place (Crispen & Hoffman, 2016). In the various literatures in which the concept of expertise is referenced, including human factors and cognitive systems engineering, we still see quite often that the method that is most often used to determine that an individual is an expert is the so-called ten-year rule of thumb. This rule originated in the work of Herbert Simon and Kevin Gilmartin (1973) on the development of mastery at chess, and an examination of the careers of famous musicians by John R. Hayes (1985), who found that masterful works were created only after about ten years of intensive effort. Since Hayes' study, a number of technical writings in the cognitive and computational sciences, and articles in the popular press, have not only mythologized the rule but have laid claim to it (e.g., Gladwell, 2008). Thus, it has become possible for researchers—even researchers who should know better—to assert that their research participants were experts because simply they had been doing their job for at least ten years (sometimes fewer). This crops up in studies in which "experts" conduct usability analysis, studies in which "experts" are the participants in human factors experiments on human-machine performance, and studies in which "experts" are the participants in knowledge elicitation.

Obviously, the ten-year criterion is insufficient. For one thing, practice alone is not sufficient since a person can progress to the journeyman level of proficiency and stay there (LaDue,

et al., 2019). Also, there are measurable individual differences in terms of how much, and what kinds of practice are required for individuals to achieve expertise (Macnamara, Hambrick & Oswald, 2014). This article elaborates a method for identifying experts.

IDENTIFYING EXPERTS

At a theoretical or conceptual level, experts have been defined by reference to the concepts of the Craft Guilds of the Middle Ages, which distinguished a number of levels of proficiency (novice, beginner, apprentice, journeyman, expert, master) (Hoffman, 1998). The Craft Guild scheme reminds us that humanity does not neatly bifurcate into people who are novices versus people who are experts—this being another myth that appears in both the popular press and (unfortunately) in the technical literatures. Additionally, the Craft Guild scheme provides a good conceptual definition of expertise: “The expert is a distinguished or brilliant journeyman, highly regarded by peers, whose judgments are uncommonly accurate and reliable, whose performance shows consummate skill and economy of effort, and who can deal effectively with certain types of rare or ‘tough’ cases. Also, an expert is one who has special skills or knowledge derived from extensive experience with subdomains” (Hoffman, 1998, p.85). This definition of expertise alludes to measures of peer review, professional judgement, experience, performance, skill, and knowledge.

Best practice in experimental psychology mandates the reliance on more than one measure for any given theoretical concept. (This is another problem with the sole reliance on the ten-year rule.) A proficiency scale for a given domain should be based on more than one of the general classes of measures, and the associated measurement methods. Five classes of methods are described in Table 1.

Table 1. Classes of methods that can contribute data for the creation of a proficiency scale. Examples are taken from studies of experts in the domain of weather forecasting (Hoffman, et al., 2017; Hoffman, Coffey & Ford, 2006; LaDue, et al., 2019) and studies in the domain of electrical utilities (Hoffman & Hanes, 2003).

METHOD	YIELD	EXAMPLE
In-depth career interviews about education, training, etc.	Ideas about breadth and depth of experience; Estimate of hours of experience and the actual primary domain tasks (versus hours on the job).	Weather forecasting in the armed services, for instance, involves duty assignments having regular hours and regular job or task assignments that can be tracked across entire careers. Amount of time spent at actual forecasting or forecasting-related tasks can be estimated with some confidence.
Professional achievements, standards, or licensing	Criteria about what it takes for individuals to reach the top of their field.	The study of weather forecasters involved individuals who had qualified to issue forecasts, including senior meteorologists US National Atmospheric and Oceanographic Administration and the National Weather Service. One participant was one of the forecasters for Space Shuttle launches; another was one of the designers of the first weather satellites.
Measures of performance at the familiar tasks	Can be used for convergence on scales determined by other methods. One should never assume that the ostensive primary task is the task at which the individual is expert. Furthermore, one should never assume that performance-based proficiency scaling should be based on performance on a single task.	Weather forecasting is again a case in point since records can show for each forecaster the relation between their forecasts and the actual weather. In fact, this is routinely tracked in forecasting offices by the measurement of "forecast skill scores."
Social Interaction Analysis (Sociometry)	Who talks to whom? Who goes to Whom for particular problems? Proficiency levels in some group of practitioners or within some community of practice (Mieg, 2000; Stein, 1997)	In a project on knowledge preservation for the electric power utilities, experts at particular jobs (e.g., maintenance and repair of large turbines, monitoring and control of nuclear chemical reactions, etc.) were readily identified by plant managers, trainers, and engineers. The individuals identified as experts had been performing their jobs for years and were known among company personnel as "the" person in their

		specialization: <i>"If there was that kind of problem I'd go to Ted. He's the turbine guy."</i>
Cognitive Task Analysis	Models of knowledge, strategies	Examples would include all the applications of the Critical Decision Method, and the projects involving knowledge modeling using Concept Maps. Models can be compared for concordance across Experts.

Based on this set of classes, I propose The Pentapod Principle: *Always use at least two and ideally three methods from the distinct methods classes to converge upon and validate a proficiency scale that is appropriate to the given domain.*

EXAMPLE: WEATHER FORECASTING

The studies of proficiency in the domain of weather forecasting illustrate the multi-method approach. In-depth career interviews were relied on the career records of civilian and military weather forecasters. It was possible to describe the depth and diversity of forecaster training and experience, and also estimate the amount of time that had been spent at actual forecasting tasks on work shifts. This included determination of the amount of time it took to qualify as a forecaster (that is, allowed to issue official forecasts). Another method was performance analysis. Forecasts are routinely evaluated post hoc in terms of what is (somewhat misleadingly) called a “skill score.” This is the value added by a forecast over and above the accuracy that would derive from a forecast based solely on climatological data. Finally, knowledge was evaluated by having the forecasters engage in Concept Mapping of their domain’s concepts, principles and atmospheric dynamics. The propositions in the knowledge models were cross-validated by having an experienced forecaster review the Concept Maps proposition by proposition.

As the data from these measures showed, and as one would hope and expect, the individuals who were identified as experts:

- Had more diverse experiences (e.g., forecasting at diverse locations having differing climates and weather tendencies),
- Knew more about domains concepts and principles, with about 90% of the knowledge propositions cross-validating (disagreements mostly involve wordsmithing),
- Were identified in social network analysis as "go-to" persons for special skill at particular forecasting problems, (e.g., hurricane tracks),

- Had spent more time at actual forecasting tasks (in some cases, well over 10,000 hours),
- Showed reliably superior performance (e.g., accuracy of 85% on the difficult task of forecasting summertime thunderstorms),
- Had developed forecasting procedures that were more refined and seasonally-dependent than those of apprentices and journeymen (who tend to over-rely on the outputs of the computer models).

Additionally, the data led to the conclusion that it is valuable, and not merely possible, to distinguish grades within levels of proficiency (e.g., junior journeyman, journeyman, senior journeyman, etc.).

CONCLUSION

The easy assumption that 10,000 hours (or ten years) experience is enough to qualify a person as expert, and the equally flawed assumption that humanity neatly bifurcates into novices versus experts—these are assumptions that feed the “war on expertise” (Klein, Shneiderman, Hoffman & Wears, 2019). This war seems to persist, especially in the popular press but also in the technical literatures. Certain claims need to be countered and disavowed, claims such as “people are surprised by the limitations in their understanding” (Fischer & Keilm, 2016, p. 1251) that are asserted in studies that are ostensibly about experts, but actually are about college freshmen who are subjects in laboratory experiments, and whose only claim to expertise is that they were “familiar” with the problem domain.

In the field of knowledge elicitation, there are instances of studies that used a multi-method approach and were arguably successful (see Table 1, above). There are numerous cases where a single “hours or years” rule was used but the consequent claim to have clearly bifurcated experts versus novices remained dubious, or at least arguable. What is lacking, and would be interesting, are cases of failure using a multi-method approach. While those may be impossible to find, it will be important in carrying out the multimethod approach for there to be criteria for evaluating its success.

References

Crispen, P. & Hoffman, R. R. (2016). How many experts? *IEEE Intelligent Systems*, 31 (4), 57-62.

- Simon, H.A. & Gilmarin, K. (1973). A simulation of memory for chess positions," *Cognitive Psychology*, 5 (1), 29-46.
- Hayes, J.R. (1985). Three problems in teaching general skills. In S.F. Chipman (et al.) (Eds.), *Thinking and Learning Skills, Vol. 2: Research and Open Questions* (pp. 391–405). Mahwah, NJ: Erlbaum.
- Gladwell, M. *Outliers*. (2008.) New York: Little, Brown and Company.
- LaDue, D.S., Daipha, P., Pliske, R.M., & Hoffman, R.R. (2019). Expertise in weather forecasting. In P. Ward, et al. (Eds.), *Oxford Handbook of Expertise*. [DOI: 10.1093/oxfordhb/9780198795872.013.38]
- Hambrick, D.Z. (2014). Malcolm Gladwell's 10,000 Hour Rule for deliberate practice is wrong. [http://www.slate.com/articles/health_and_science/science/2014/]
- Macnamara, B.N., Hambrick, D.Z., & Oswald, S.J. (2014). Deliberate practice and performance in music, games, sports, education and professions: A meta-analysis. *Psychological Science*, 25 (8), 1608-1618.
- Hoffman, R.R. (1998). How can expertise be defined?: Implications of research from cognitive psychology. In R. Williams, W. Faulkner and J. Fleck (Eds.), *Exploring expertise* (pp. 81-100). New York: Macmillan,
- Hoffman, R.R., LaDue, D., Mogil, H.M., Roebber, P., & Trafton, J.G. (2017). *Minding the Weather: How Expert Forecasters Think*. Cambridge, MA: MIT Press.
- Hoffman, R.R. & Hanes, L.F. (2003). The boiled frog problem. *IEEE Intelligent Systems*, 18 (4), 68-71.
- Hoffman, R.R., Coffey, J.W., Ford, K.M. & Novak, J.D. (2006). A method for eliciting, preserving, and sharing the knowledge of forecasters. *Weather and Forecasting*, 21 (3), 416–428.
- Klein, G., Shneiderman, B., Hoffman, R.R. & Wears, R.L. (2019) The war on expertise." In P. Ward (et al.) (Eds.), *The Oxford Handbook of Expertise*. [DOI: 10.1093/oxfordhb/9780198795872.013.50]
- Fisher, M. & Keil, F.C. (2016). The curse of expertise: When more knowledge leads to miscalibrated explanatory insight. "*Cognitive Science*, 40 (5) p. 1251-1269.