Who Should Design for Novice Users? A Study of Experts' Limitations in Predicting Novices' Experience

Pamela Hinds
Integrated Solutions Laboratory
HPL-98-136
August, 1998

expertise, human-computer interaction, cognitive biases, estimation accuracy

Experts are often called upon to design interfaces, applications, and instruments for novices. But, there are reasons to believe that expertise may inhibit peoples' ability to understand the novice experience. Data from one study suggests that experts suffer from a cognitive bias that leads them to underestimate the difficulty that novices will encounter when performing an unfamiliar task. The data also suggest that experts' bias toward underestimation is resistant to a commonly used debiasing technique – listing the steps required to perform a task and listing some of the problems that could occur when novices attempt a task for the first time. These results reinforce the importance of including multiple perspectives in the design process and making sure that expert designers work to understand the perspective of their novice users. The results also suggest that simply reading usability reports on the problems experienced by novices may not be adequate to debias experts. It may be necessary to create better ways to facilitate experts' understanding of novices' experience.

Internal Accession Date Only

Introduction

Experts are often called upon to design systems for novices. But, even with tremendous effort and dedication, designers inadvertently build systems that users find difficult to use. In fact, as many as ¾ of all large systems either don't function as intended or are never used (Science, 1994). Although there are many aspects to the usability issue, one question to ask is whether or not we have the right people involved in designing systems for novices. Are experts the right people to be developing systems for novice users or are experts biased by their own experiences? If experts do exhibit some bias in being able to anticipate the experience of novices, it also is important to understand what generates that bias and to develop tools to counteract the bias. The research reported here asks two questions:

How good are experts at estimating novices' experience?

What debiasing tools can be used to improve accuracy in experts' estimates?

Extensive research in cognitive science has examined the differences between experts and novices. This research shows that experts are able to detect meaningful patterns more rapidly than novices. Experts also tend to evaluate information based on a set of principles (Soloway, Adelson, & Ehrlich, 1988), spend more time in analysis (Paige & Simon, 1966), organize information by collapsing task elements into larger groupings (Langer & Imber, 1979), and perform tasks more quickly than novices (Anderson, 1982). This body of research establishes that experts are more thorough, more principled, and more accurate in performing tasks within their domain of expertise.

Other research suggests that there are significant limitations to expertise. In judgment tasks, experts perform no better than novices and often are less accurate than mathematical (regression) models (Dawes, 1971; Johnson, 1980). For example, when predicting the performance of new medical interns and residents, a panel of experts were only slightly more accurate than novices and significantly less accurate than a regression model (Johnson, 1980).

When estimating performance, people generally underestimate their own and others' performance (Buehler, Gruffin, & Ross, 1994). And, there are several reasons to believe that experts may be more biased toward underestimation than non-experts. First, as people gain expertise, they begin to reorganize their view of the task into fewer categories (Langer & Imber, 1979). This collapsing of elements into categories may serve to simplify the task in the minds of experts and lead to underestimation. Second, in many tasks, people tend to use an anchoring and adjustment heuristic (Nickerson, Baddeley, & Freeman, 1987). Anchoring and adjustment is a process whereby people anchor on some value and then adjust that value to attain an estimate. Research shows that the adjustment process often falls short of bridging the gap between the anchor and "reality". In the case of estimating novices' performance, experts may anchor on their own more rapid, trouble-free performance of the task and inadequately adjust for the novices' lack of knowledge and experience. Third, people who have become experts are likely to have developed this expertise over time. Time and practice may make experts' experience as novices less available (Kahneman & Tversky, 1979). Taken together, these findings suggest that expertise may lead to increased underestimation of novice task completion times.

> *Hypothesis 1:*
> *People with more expertise as compared to those with less expertise will be less accurate estimators of novices' task completion times, with a bias toward underestimation.*

Research on debiasing has demonstrated that people recall events more accurately when they create a list of activities to trigger their memories of the task (Engle & Lumpkin, 1992). If experts are prone to debiasing because they have simplified the task and forgotten their own experience as novices, then a list may improve experts' estimates of novice performance times.

> *Hypothesis 2:*
> *Providing a list of task elements and potential problems will reduce experts' bias toward underestimation.*

Method

The study was a 2 X 2 between subjects experiment with expertise (low, high) and debiasing (unaided, list) as factors. Participants were asked to build a LEGO toy and to predict how long it would take the average student to build the same toy the first time. The LEGO toy was a 39-piece "V-Wing Fighter" (see figure 1). Participants were given two pictures of the assembled toy–one side view and one front view–but were not given step-by-step instructions for assembling the toy.

Participants

Participants were 49 male undergraduate students in non-technical disciplines (e.g. business, art, economics, psychology, etc.) and were recruited through electronic and physical bulletin boards at Carnegie Mellon University.

Expertise Manipulation

The development of expertise relies on extensive practice and usually develops over time (Chase & Ericsson, 1982; Chase & Simon, 1973; Ericcson & Charness, 1997). In this study, participants attended two sessions. In the first session, those in the high expertise condition completed a pre-test survey and assembled the LEGO toy four times before the experimenter dismissed them. Participants in the low expertise condition completed the pre-test survey and were dismissed without building the LEGO toy. The second session was scheduled 10 to 14 days after the first session. In the second session, all participants built the LEGO toy one time and were then asked to

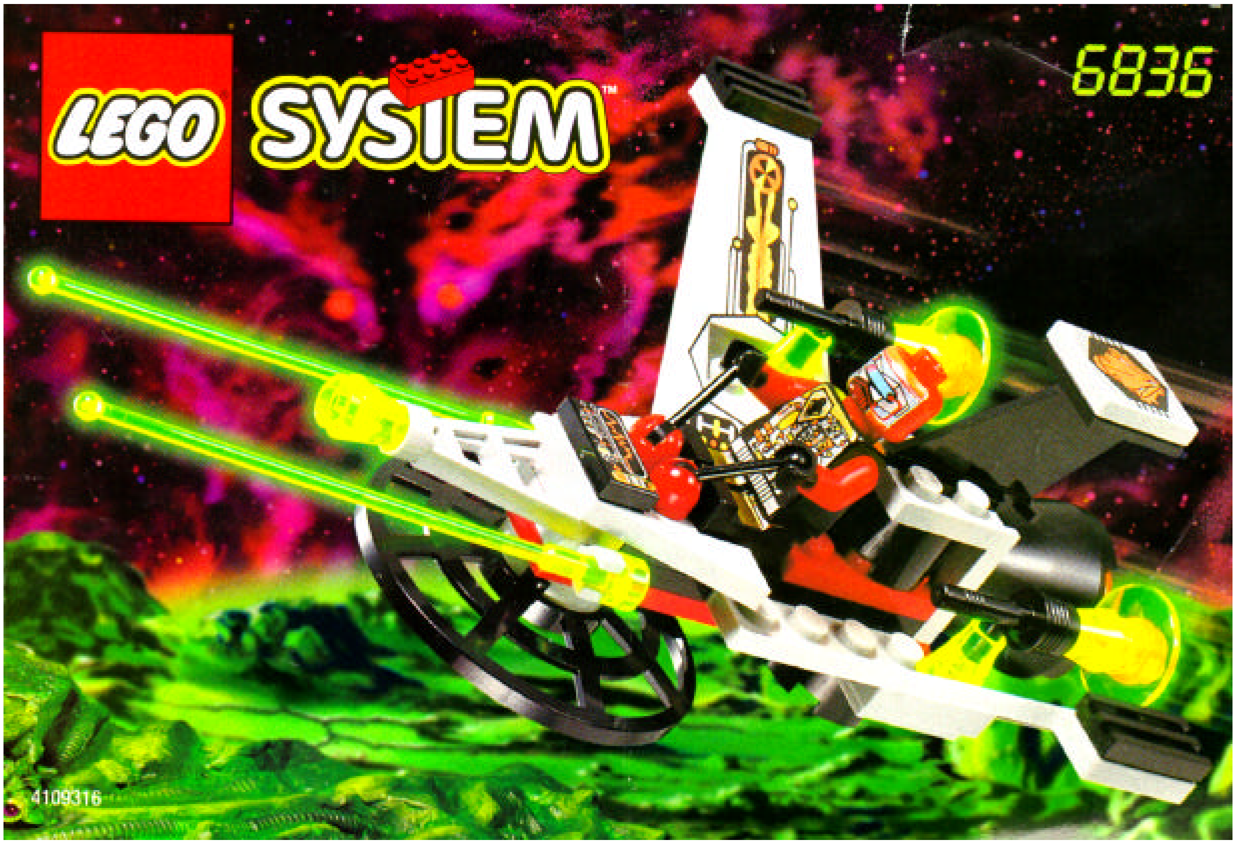

Figure 1: LEGO Toy assembled by participants in study.

provide a time estimate of how long it would take novices to complete the task. Therefore, when making their prediction of novices' performance times, those in the high expertise condition had built the toy five times and those in the low expertise condition had build the toy only once. Figure 2 illustrates the order of activities for each condition in the study.

Debiasing Manipulation

Participants were asked to make their predictions based on either the unaided (control) or the list elicitation method. In all conditions, participants were told:

> *We are planning a future experiment and need an estimate of how long it will take people to construct this same toy for the first time. These people will be undergraduate students at CMU in non-technical majors (i.e., industrial management, literature, etc.) and will be constructing the toy in an environment similar to this one. They will be given instructions and pictures from which to work that are identical to those given to you.*

Condition	Session #1	Session #2
Low expertise (Novice)	• complete pre-test	• build toy 1 time • provide estimate • complete post-test
High expertise (Expert)	• complete pre-test • build toy 4 times	• build toy 1 time • provide estimate • complete post-test

Figure 2: Sequence of activities in study.

The unaided elicitation simply asked participants to provide an estimate of "the median time it would take for a person like the one described above to build this toy for the first time." The list debiasing method provided a list of activities required to build the toy (i.e. figuring out how to connect the LEGO pieces) and problems novices encountered in the pre-test (i.e. differentiating pieces that look the same). Participants in the list debiasing conditions were asked to "write out a scenario of what a person's experience might be like, given the potential problems listed above" and then to give a median estimate in minutes.

<u>Procedure</u>

All participants attended two session between 10 and 14 days apart. In session one, participants arrived at the laboratory and were asked to complete a pre-test that included demographic questions and questions about their experience with LEGO toys. If they were randomly assigned to the low expertise condition, participants completed the pre-test, confirmed their second appointment, were paid for their participation at the first session, and were dismissed. Participants randomly assigned to the high expertise condition completed the pre-test and were then instructed to build the toy. When they were finished building the toy, the experimenter verified that the toy had been built correctly, gave the participant another unassembled V-Wing Fighter and instructed the participant to build the toy again. After the participant built the toy four times, the experimenter confirmed the second appointment, paid the participant for the first session, and dismissed them.

In the second session, participants arrived in the same laboratory and all were instructed to build the toy. After the participants built the toy once, they were asked to provide an estimate of the median time it would take a novice to complete the task under similar conditions. They then completed a post-test survey that included questions about the participant's confidence in their estimate, the process they used to generate the estimate, the steps in the LEGO task, and their own performance time(s). After completing the post-test, participants were debriefed, paid for their participation in the second session, and dismissed.

Analysis
The primary dependent variable was predictive accuracy–participant's estimate minus the median actual task completion time of novices in the study. Median novice completion time was calculated using the first performance time of participants in all conditions in the study. ANOVA were used to determine the relationship between expertise, debiasing method, and predictive accuracy.

Results
Novice, on average, took 12.5 (SD=5.08) minutes to complete the task. The range was from 3 to 27 minutes with a median of 12.27 minutes. Research participants, on average across conditions, predicted it would take only 10.7 (SD=5.68) minutes for novices to complete the task. These data are consistent with other research identifying a general bias toward underestimation. When performing the task for the first time, there was no significant difference in the performance times of those in the high versus the low expertise conditions, $F[1, 46]=.09$, ns., with those in the high expertise condition taking 12.7 (SD=5.77) minutes on average and those in the low expertise condition taking 12.3 (SD=4.40) minutes on average.

Table 1 shows the mean estimates for all conditions. Those in the high expertise condition, on average, predicted that it would take 8.5 (SD=3.36) minutes for a novice to perform the task. In contrast, those in the low expertise condition predicted that it would take 12.9 (SD=6.73) minutes for a novice to perform the task. Whereas those high in expertise underestimated novice performance time by nearly 4 minutes, those low in expertise overestimated novice performance time by about 35 seconds.

Table 1
Mean Estimate (Standard Deviation) in Minutes by Debiasing Condition and Level of Expertise

	Debiasing Condition	
Level of Expertise	Unaided (N=26)	List (N=23)
Low	11.46 (5.44)	14.64 (7.92)
High	8.31 (2.29)	8.75 (4.33)

Note. Actual median novice time was 12.27 minutes.

Hypothesis 1 posited that experts predicting novice task completion times would be more biased toward underestimation than would be those with less expertise and hypothesis 2 posited that providing a list of task elements would reduce this bias. To test these hypotheses, a 2-factor ANOVA was performed with expertise (low and high) and debiasing condition (unaided and list) predicting accuracy. Accuracy was calculated by subtracting participants' estimates from the actual median performance time of novices in this study. The ANOVA results indicate a significant difference in prediction by level of expertise, $\underline{F}(1, 45)=8.96$, $\underline{p} <.01$. As in study 1, those with more expertise more severely underestimated novice performance times.

The data from table 1 suggest that people low in expertise improved when given the list debiasing method, but that people high in expertise were resistant to debiasing. Although those in the low expertise condition increased their estimates by over 3 minutes when given the debiasing list, those in the high expertise condition increased their estimate by less than 30 seconds. The ANOVA results indicate no main effect for debiasing, $\underline{F}(1,45)=1.44$, ns., and no interaction effect between expertise and debiasing, $\underline{F}(1,45)=.82$, ns., suggesting that neither group was particularly responsive to the list debiasing method. Because they were so close in their estimate without debiasing, the list reduced predictive accuracy of those low in expertise. Those in the unaided condition underestimated less than one minute (at 11.5 minutes), whereas those in the list debiasing condition overestimated by over 2.5 minutes (at 14.64 minutes).

Discussion

Although it might be expected that experts' superior knowledge and experience should lead them to be better predictors of novice task completion time compared to those with less expertise, the findings in this study suggest otherwise. More specifically, the results reported here suggest that experts' superior knowledge actually interferes with their ability to predict novice task performance times.

Not only were experts unable to take advantage of their knowledge and experience, they were also unable to correct their estimates when they were prompted with a presentation intended to help them reduce their underestimation. The results from this study shows that experts have difficulty to incorporating information about problems that novices have encountered. The inaccuracy of experts and their resistance to prompting has a number of possible explanations.

Evidence suggests heavy reliance on an anchoring and adjustment heuristic with experts' own novice performance as the reference point. But, those at both levels of expertise described a similar anchoring strategy when asked to describe their estimation process. The primary difference between those with more and those with less expertise was the accuracy with which they recalled their own novice performance of the task. Experts underestimated their own novice performance times by over 3 minutes and they underestimated other novices' performance times by nearly 4 minutes. Experts' bias toward underestimating their own novice performance, combined with an anchoring process that relies on accurately recalling their novice performance (or making an adequate adjustment), nearly guaranteed experts' underestimation of others' novice performance times. Evidence from these data suggests that the availability bias is the key contributor to experts' relative inaccuracy in estimating novice performance times.

Experts were also resistant to debiasing. There are several possible explanations for their resistance. First, experts may have difficulty understanding the challenges faced by novices, even when reminded of these challenges. Even when provided a list of problems, experts might be unable to understand novices' difficulties in identifying the problem, finding a solution, and implementing that solution. Such an explanation is consistent with the "curse of knowledge," a bias in which knowledgeable people are unable to ignore their own superior knowledge in trying to make naive predictions (Camerer, Loewenstein, & Weber, 1989).

Second, experts may be unduly confident in their original estimates. Experts have been found to be overconfident in judgments related to their field of expertise and to become increasingly confident, although not more accurate, with increasing information (Oskamp, 1965). In fact, less than 15% of the experts in Oskamp's study changed their answers when given additional pertinent information. Experts may have been overconfident in their original predictions of novice task performance times and been unwilling to reconsider their estimates when presented with the recall and the list methods. But, data from this study do not support this line of reasoning. When asked about their level of confidence in their estimate, those high in expertise reported marginally lower levels of confidence than those low in expertise, $F(1, 47)=3.31$, $p < .10$.[1] Therefore, overconfidence alone is an unlikely explanation for experts' resistance to debiasing in these studies.

Third, there are a multitude of possible debiasing tools that could be developed. In the research reported here, I examine only two. It is possible that there are other debiasing methods that could improve experts' accuracy. The results suggest that experts engage in an anchoring and adjustment process in which their anchor is biased by the unavailability of their own novice experience. This finding suggests several possible directions for debiasing tools. One direction is to prompt experts to use a different anchor, possibly suggesting that they think about a task they have recently learned or watched someone else learn. Another direction is to improve experts' recall of their own novice experience. The list method was intended to accomplish this, but a stronger manipulation may be required to jog experts' memories. Another approach may be to increase people's motivation for accuracy. For instance, increasing peoples' accountability for their judgments has been found to significantly reduce cognitive biases (Tetlock, 1992).

Although future research on debiasing tools is important, it may be that the expert bias runs too deep and will remain resistant to debiasing methods. If that is the case, there

[1] Consistent with other research, there was no relationship between confidence and predictive accuracy, $F(1.47)=.82$, ns., in study 2.

may still be several alternatives for improving experts' estimates of novice performance times. One obvious alternative is to build expertise in predicting novice performance. As experts receive feedback on the accuracy of their estimates, they are likely to improve. Some organizations build such expertise in management recruiters, marketeers, usability engineers and other professionals. Another alternative is to use intermediates to assist in predicting novice performance. I would not suggest eliminating the expert from this process, but closing the gap by adding in the perspective of someone closer to the experience of the novices being predicted. Finally, having experts observe novices may help experts to understand the problems encountered by novices. This is a common practice in human-factors laboratories, although anecdotal evidence of its effectiveness at reducing experts' bias is mixed.

Although these study results provide strong evidence of experts’ bias toward underestimation, several limitations must be addressed. First, true expertise is a condition that develops over time with extensive practice and may be difficult to create in a laboratory experiment. Some have argued that 100 hours or more of practice over 10 or more years are required to become expert in cognitive tasks (Anderson, 1982; Ericsson and Charness, 1994). But, as Sternberg (1997) has argued, people are not expert or non-expert, but reside somewhere along a continuum of expertise. In this study, a relatively simple task was selected so that some degree of expertise could be acquired quickly. Participants in the high expertise condition performed the task a total of 5 times. These participants may not be experts in the pure sense of the term, but they have acquired enough expertise to outperform novices and to have reached a plateau of performance. Experts, on average, took less than 6 minutes to build the LEGO toy the fifth time they built it compared to nearly 13 minutes the first time. And, their performance began to plateau at the third build at when they completed it in 3.9 (SD=1.05) minutes, on average.

Second, researchers have suggested that experts' performance is sometimes tested with tasks outside the domain of the experts' experience (e.g. Bolger & Wright, 1992). One could argue that the participants in these studies were not expert at estimating task completion times of novices and that given adequate experience in estimating novices' performance, they would become more accurate. Future research could explore how experts' experience in estimating others effects their accuracy. Research on experts' ability to learn from different kinds of feedback also would be valuable. Another limitation is that only two tasks were examined in these studies and both tasks required that participants interact with a physical object(s). It is possible that experts may respond differently with more cognitively demanding and more cognitively complex tasks. Replication with tasks in other domains and with more complexity is necessary to determine the generalizeability of these results. Third, although participants in study 1 were relatively expert in cellular telephony, they would not have been internationally renowned as experts in the domain. It is possible that exceptionally talented people within a domain are less likely to fall prey to this bias. Future research with participants elsewhere on the continuum of expertise is warranted to fully understand the shape of the relationship between expertise and predictive accuracy.

These findings have implications in a variety of settings where people are predicting the performance of those new to an organization, product, or task. The results suggest that project managers may not be the most accurate at estimating project completion times for novice team members. Marketers may not be the most appropriate people to anticipate the difficulties consumers' face when learning to use a new product. And, designers may not be the best estimators of what novices' experience will be when trying to learn to use a new piece of technology.

These data also suggest that haranguing experts to improve their predictions of novice performance may not be productive. If experts have a cognitive handicap when it comes to predicting novices' experience, then applying more pressure may only cause frustration. Identifying effective debiasing methods or structuring the design process so that experts' bias is balanced with a different perspective may result in more accurate time estimates and more usable products.

References

Anderson, J. R. (1982). Acquisition of cognitive skill. Psychological Review, 89, 369-406.

Buehler, R., Griffin, D, & Ross, M. (1994) Exploring the "planning fallacy": Why people underestimate their task completion times. Journal of Personality and Social Psychology, 67, 366-381.

Camerer, C., Loewenstein, G. & Weber, M. (1989). The curse of knowledge in economic settings: An experimental analysis. Journal of Political Economy , 97, 1232-1254.

Chase, W. G. & Ericsson, K. A. (1982). Skill and working memory. In G. H. Bower (Ed.), The psychology of learning and motivation (Vol. 16, pp. 1-58). New York: Academic Press.

Chase, W. G. & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), Visual information processing (pp. 215-281). New York: Academic Press.

Dawes, R. (1971). A case study of graduate admissions: Application of three principles of human decision making. American Psychologist, 26, 180-188.

Engle P. L. & Lumpkin, J. B. (1992). How accurate are time-use reports? Effects of cognitive enhancement and cultural differences on recall accuracy. Applied Cognitive Psychology, 6, 141-159.

Ericsson, K. A. & Charness, N. (1994). Expert performance: Its structure and acquisition. American Psychologist, 49, 725-747.

Ericsson, K. A. & Charness, N. (1997). Cognitive and developmental factors in expert performance. In P. J. Feltovich, K. M. Ford, & R. R. Hoffman (Eds.), Expertise in Context: Human and Machine (p. 3-41). Menlo Park, CA: AAAI Press.

Johnson, E. J. (1980). Expertise in admissions judgment. Unpublished Doctoral Dissertation, Carnegie-Mellon University, Department of Psychology.

Kahneman, D., & Tversky, A. (1979). Intuitive prediction: Biases and corrective procedures. TIMS Studies in Management Science, 12, 313, 327.

Langer, E. J. & Imber, L. G. (1979). When practice makes imperfect: Debilitating effects of overlearning. Journal of Personality and Social Psychology, 37, 2014-2024.

Nickerson, R. S., Baddeley, A., & Freeman, B. (1987). Are people's estimates of what other people know influenced by what they themselves know? Acta Psychologica, 64, 245-259.

Oskamp, S. (1965). Overconfidence in case-study judgments. Journal of Consulting Psychology, 29, 261-265.

Paige, J. M. & Simon, H. A. (1966). Cognition processes in solving algebra word problems. In B. Kleinmuntz (Ed.), Problem Solving (p. 119-151). New York: Wiley.

Soloway, E., Adelson, B., & Ehrlich, K. (1998). Knowledge and processes in the comprehension of computer programs. In M. Chi, R. Glaser, & M. Farr (Eds.), The Nature of Expertise (p. 129-152). Hillsdale, NJ: Lawrence Erlbaum.

Sternberg, R. J. (1997). Cognitive conceptions of expertise. In P. J. Feltovich, K. M. Ford, & R. R. Hoffman (Eds.), Expertise in Context: Human and Machine (p. 149-162). Menlo Park, CA: AAAI Press.

Tetlock, P. E. (1992). The impact of accountability on judgment and choice: Toward a social contingency model. In M. P. Zanna (Ed.), Advances in experimental social psychology (Vol. 25, pp. 331-376). New York: Academic Press.