Diversity and Software Development

Hakan Erdogmus

We’re told that diversity is good for society. And we need to be told, because if we’re left to our own instinctive devices, we’d probably avoid it: human nature tends to favor the attraction of likes. But do we know why diversity is good in the first place? According to Scott E. Page, the University of Michigan professor of complex systems, political science, and economics and the author of *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies* (Princeton University Press, 2007), diversity is good beyond altruistic and moral reasons. And many insights drawn from his thinking apply to software development practice, research, and education.

Page suggests that diversity enhances our ability to tackle hard problems and make accurate predictions. Our ability to tackle hard problems affects our productivity, and our ability to make good predictions impacts how we make decisions and deal with risk and uncertainty. Clearly, software development can benefit from improvement in both dimensions.

Page’s conclusions aren’t based on an empirical model. His theoretical framework builds on a few simple but beautifully intuitive concepts, definitions, and axioms. His insights are based on provable mathematical assertions with sufficient and necessary conditions rather than on rules of thumb based on observations of human behavior. Thus, they are refutable to the extent that the underlying concepts, definitions, and axioms fail to capture real-world behavior or to the extent that the stipulated conditions fail to hold true. And surely, such failure is possible in many situations. The insights are nevertheless invaluable.

What Does Diversity Mean?

Page reduces diversity to differences in the sets of intellectual tools that people use in problem solving and decision making. He classifies these tools into perspectives, interpretations, heuristics, and predictive models. Perspectives are ways of representing the world. Related to perspectives, interpretations are ways of abstracting the world; in my mind, they’re similar to Peter M. Senge’s mental models (*The Fifth Discipline: The Art and Practice of the Learning Organization*, Doubleday Currency, 1990) that limit us to particular ways of thinking. Interpretations are the internal filters, projections, and categorizations we use on top of perspectives. Heuristics are ways of exploring a solution space. Finally, predictive models are ways of inferring causality and using causality to guess outcomes.

Page maintains that the number and combination of these intellectual tools dictate how good people and groups are in finding solutions and making predictions, depending on the task at hand. The larger a group’s collective toolbox is and the more variety it features, the better the outcomes are. Said otherwise, the results get better as a group gets more diverse. Hence, diversity implies both the quantity and variety of the available tools.

Mission Statement: To be the best source of reliable, useful, peer-reviewed information for leading software practitioners—the developers and managers who want to keep up with rapid technology change.
Diversity in Problem Solving

Diversity isn’t useful in all problem-solving situations: it might even be counterproductive. For example, diversity doesn’t matter in menial or mechanical tasks. For hard problems, though, it should matter. And software development has plenty of those.

Page explains that diverse perspectives and heuristics are central to solving hard problems. Diverse perspectives increase the number of solutions that a group may collectively generate by exploiting connections among the individual pieces of a puzzle being solved. Diverse heuristics let problem solvers explore a larger portion of the solution space without getting stuck. Then unsurprisingly, everything else being equal, diversity beats homogeneity.

However, what’s much less obvious is that diversity also beats ability, albeit under certain conditions. Page’s research proves that given a hard-enough problem, a random group of smart people capable of contributing to that problem’s solution will on average outnumber the best individual problem solver that problem’s solution will on average. Diverse heuristics let problem solvers explore a larger portion of the solution space without getting stuck. Then unsurprisingly, everything else being equal, diversity beats homogeneity.

This result is remarkable and has immediate implications for software organizations and teams. We can safely assume that developing software of nontrivial functionality qualifies as a hard problem, at least in most cases. The first implication of the “diversity beats ability” principle is to draw the talent to compose the team from a broad pool that covers a variety of backgrounds, skills, and experience. The second implication is to focus efforts on mixing and matching those backgrounds, skills, and experiences to form the team rather than on identifying the few elitist stars. The second implication in particular flies in the face of the popular anecdote that people-focused development practices work only when the team comprises the best people. Note that Page’s conjecture doesn’t mean that you shouldn’t find the best people you can. It only means that a diverse group of average yet sufficiently competent people who collectively possess numerous perspectives and heuristics might do as well or better than a homogeneous group of all rock stars with overlapping perspectives and heuristics.

Diversity in Prediction

In Blink: The Power of Thinking without Thinking (Little, Brown and Company, 2005), Malcolm Gladwell describes how experts learn to look at just a few features of a problem and make amazing predictions. Page, however, cautions us that in some situations even the best “blinker” won’t do well. I’m afraid that the highly multifactorial and uncertain world of software estimation—in fact, most decision making that takes place in the context of software development—falls under the “not-blinkable” umbrella.

Page provides three main insights regarding prediction. First, diversity of interpretations and predictive models is central. Second, ability and diversity are equally important—in his words, “being different is as important as being good.” And finally, crowds outperform averages in predictive tasks: a diverse group’s collective estimate is on average more accurate than the average accuracy of the individual estimates in that group. The latter two conjectures partially explain the compelling famous anecdotes James Surowiecki recounts in The Wisdom of Crowds (Random House, 2004).

Experts or Crowds?

Should we then go with experts or crowds in predictive software development tasks? Page’s two prediction conjectures don’t address this question directly. The answer is more complicated than we’d like it to be. Page demonstrates that a really good expert will on average perform better than a not-so-wise group, but not necessarily in all cases. For a common prediction task, a small chance that an ordinary group outperforms an expert who is far more able than the group’s individual members may give rise to a large number of anecdotes, and in turn, a persuasive story. As persuasive as the story might be, it will still not override the truth.

So Page’s general advice is to go with the expert only if the person is known to be far more accurate than a group’s individual members: it wouldn’t be enough for the expert to be just moderately better.

Experts or Regression Models?

Page further draws attention to the bulk of evidence in economics, political science, and other disciplines regarding the superiority of regression models based on historical data over judgment-based expert predictions. In software estimation,
which approach performs better is still a topic of intense debate, as exemplified by the Viewpoints piece “Software Development Effort Estimation: Formal Models or Expert Judgment?” by moderator Stan Rifkin and debaters Magne Jørgensen and Barry Boehm (IEEE Software, March/April 2009). Experts rely on a few variables at a time in making predictions. Page states that when the prediction task becomes difficult in a complex multivariate world, a single expert’s prediction may not be much better than random guessing.

So why then should we bother with experts and not just use regression models? Page is right on the mark when he suggests that we actually do use experts in any case. Experts are the ones who determine which variables matter and how important they are. They are the ones who collect, organize, and interpret the data, construct and calibrate the regression models, and choose the parameters’ values. We also need experts when historical data isn’t available or reliable enough to be of any use.

Page’s book provides two additional findings that software estimation doesn’t orderly leverage. The first finding allows improvements upon judgment-based expert predictions: even if individual experts are inaccurate, a diverse group of experts might collectively be surprisingly accurate. The second finding allows improvements upon data-based regression models and circumvents the overfitting problem: an ensemble of simpler but orthogonal models, or the collective prediction of a set of diverse models, often outperforms a single complex model. The latter finding suggests rethinking the drive toward more and more complex estimation models.

A central tenet of Page’s theory is that accurate prediction by a diverse group doesn’t rely on iterative convergence of estimates, as does, for example, the Delphi method. On the contrary, it requires the independence of the estimators’ predictive models. Again, this point warrants rethinking existing group-based estimation approaches.

Caveats

Diversity can be undesirable in some situations. One form of potentially counterproductive diversity is what Page calls “diverse fundamental preferences.” Fundamental preferences are about individuals’ beliefs and value systems. Differences in these deeply engrained mechanisms might spoil collective decision making. When individual fundamental preferences fail to aggregate naturally, the end result is either conflict or convergence to majority rule through compromise.

Independent thinking is a prerequisite for diversity. Diversity often diminishes in a closely knit group over time as the group gradually succumbs to group-think. Page demonstrates that in such settings, random outcomes, even irrational ones, are possible as people tend to move too far in the direction of the majority opinion through peer pressure or influence. Perhaps sustained cohesiveness and success are elusive in long-lived teams because of this effect. Ironically, software processes that focus on people and collaboration sometimes devalue or ignore independent thinking. Should these processes explicitly consider the erosive effects of close interaction on diversity?

The Role of Identity Diversity

Page also touches upon identity diversity, determined by affiliation with an identifiable social group such as those based on gender, culture, race, ethnicity, religion, or sexual orientation. Identity diversity is relevant to problem solving and prediction to the extent that it leads to cognitive diversity, the kind that differentiates among individuals on the basis of the intellectual tools they possess. If identity diversity indeed leads to cognitive diversity in software development, the efforts to make information technology, computer science, and software engineering education programs attractive to underrepresented groups, minorities and women in particular, are well justified.

In the end, we still don’t know how to measure the cost-effectiveness of diversity in specific situations beyond a handful of generalized principles that emerge from Page’s and others’ research. Even if in certain contexts we find ways to gauge individual ability and the goodness of outcomes reasonably well, it isn’t clear how the economics would play out, say, when the choice is between a large, diverse team of fairly able people and a smaller but less diverse team of gurus. Now that sounds like a pretty hard problem—one fit for a diverse group of experts to attack.