Reducing Biases in Evaluation

 Ernest R. House

(Presented to New Mexico Evaluators, Albuquerque, NM, 11/15/16)

A critical task for evaluators in any evaluation is to identify potential biases that might impede their conduct of an honest, accurate, and fair evaluation. Potential biases depend on what’s being evaluated, the approach to evaluation, and the context of the study. In my first evaluation, the evaluation of the Illinois Gifted Program in 1967, the sponsors of the evaluation, the Illinois Department of Education, pressed me to make an early report to the Illinois legislature to assure legislators that the study was worthwhile. The field of evaluation was new, and the department wanted to show utility for such a novel enterprise.

To produce a quick preliminary report, we mailed questionnaires to hundreds of schools. What we got back was mostly garbage. School administrators were reluctant to respond openly to an inquiry from a group identified with state authority. For example, the questionnaires from the Chicago high schools were all filled out exactly the same way, in the same handwriting, and sent back together, with the questionnaires tied together with a ribbon. According to the responses, every Chicago high school was doing everything possible for gifted students. I can’t remember the color of the ribbon, but I got the message.

We had to rethink what we were doing and discover ways to compensate for the response bias we had underestimated. One strategy was to send well-trained, face-to-face interview teams to a sample of school districts to search for program effects. It’s one thing to fill out a questionnaire from unknown people and quite another to face well-informed interviewers collecting data in your school district.

To send teams across the state was expensive, and we decided to sample the districts and generalize from the sample. Illinois had a thousand school districts, but most students were concentrated in a few larger ones. Fortunately, experts had already sorted out the sampling bias issues. We drew a stratified random sample of districts from which to make generalizations and reported the findings to legislators in several reports.

During that project something else caught my attention-- how political evaluation was. I was forced to deal with the legislature, the education department, Chicago, hundreds of school districts, teachers, students, the staff of the state program, curriculum developers and researchers in universities, and so on. Many encounters were political. I had little preparation for the politics involved, and yet I had to sort things out. I did the best I could intuitively. Later, thinking about the interactions, I decided that an explicit conception of social justice might be useful in helping evaluators through the politics. In 1976 I published a social justice framework that might be useful in such situations.

Most evaluators now are familiar with sampling biases, response biases, and the biases discussed in textbooks. What about biases that evaluators might harbor in their beliefs and dispositions? There’s considerable evidence to indicate that personal biases can affect evaluations significantly.

In his book *Thinking, Fast and Slow*, Nobel Prize winner Daniel Kahneman (2011) has presented a model of how the human mind works. According to cognitive researchers, people make systematic errors in their thinking, and some errors are attributable to the design of their cognitive machinery. There are two basic thinking processes. System 1 is intuitive and operates automatically with little apparent effort. If you see a photo of an angry woman, you’ll recognize intuitively that she’s angry. On the other hand, if you multiply 27 times 46, you won’t know the answer, but you can solve the problem. The second process is System 2 slow thinking.

System 1, fast thinking, includes detecting that some objects are more distant than others, driving a car on an empty road, and understanding simple sentences in your native language. These abilities are learned through experience, association, and practice. Knowledge is stored in the associative memory and easily accessed. System 1 processing is mostly automatic.

System 2, slow thinking, focuses attention on problems that demand concentration, such as doing calculations, filling out tax forms, and checking the validity of arguments. “Pay attention” is the motto. These tasks require concentration and are disrupted without it. The number of System 2 tasks you can perform simultaneously is limited. You can concentrate on only a few things at a time. Most mental processing occurs in System 1, which operates while we’re awake and System 2 is in low-effort mode.

System 1 offers impressions, intuitions, and intentions to System 2. If everything seems okay, System 2 accepts these as valid. If something seems abnormal, System 2 might deal it. It searches memory for an explanatory story. However, System 2 is notoriously lazy. Only when System 1 needs help does System 2 act. Even then, it exerts minimum effort. Consequently, System 1 chooses most thoughts and actions, not System 2 (Kahneman, 2011, p. 31).

“System 1 provides the impressions that often turn into your beliefs, and is the source of the impulses that often become your choices and your actions. It offers interpretations of what happens to you and around you, linking the present with the recent past and with expectations about the near future. It contains the model of the world that instantly evaluates events as normal or surprising. It is the source of your rapid and often precise intuitive judgments. And it does this without your conscious awareness of your activities. System 1 is also…the origin of many systematic errors in your intuitions” (Kahneman, 2011, p.58).

This division of labor is efficient because System 1 is very good at what it does. Its assessments of familiar situations are swift, accurate, and appropriate because they’re based on repetition and long experience. It operates with heuristics that enable it to arrive at quick assessments. This allocation of attention was honed by evolution. Responding quickly to threats enhances survival. That’s the purpose, acting in emergencies with quick decisions. Spotting a lion that’s not there is a small price to pay for a quick response if one is there.

When jumping to conclusions, System 1’s criterion of success is the coherence of the story it pieces together. Ideas come from the associative memory and whatever the mind has been primed with. System 1 considers that which is familiar to be trustworthy, and impressions of familiarity are based on repetition. A sense of trustworthiness is often derived from sheer routine. System 1 relies on cognitive ease and lack of stress as markers of truth. It’s not sensitive to either the quality or quantity of data. Consistency is what counts. Hence, coherence leads to confidence, and sometimes to over-confidence. The potential for mistakes is apparent.

System 1 gets into trouble when the situation is unfamiliar and pattern recognition is mistaken, or if the short cuts it uses are flawed. System 1’s repertoire of short-cuts includes heavy emphasis on the familiar, judging based on few data points, assessing based on cues, relying on halo effects, categorizing by exemplars and prototypes, and interpreting through familiar frames.

What does this mean for evaluation? My interpretation is that the processes of human thought are evaluative (House, 2015). System 1 is a monitoring evaluation, and System 2 evaluates the responses of System 1. Both systems of thinking are evaluative. In other words, evaluating is a natural thought process. Evolution has shaped us in such a way that we evaluate to survive. We are evaluating organisms able to perform formidable feats of evaluation. But we also have some errors built into our thinking processes, particularly in the shortcuts enabling quick judgments.

Common biases include improper framing, such as casting evaluation criteria too narrowly; relying on incomplete evidence, such as omitting critical data from the study; seeking only data that confirm our impressions, such as halo effects and failing to look for negative instances; improper priming, such as being directed only by information from the sponsor and program director; and overconfidence about conclusions.

We should be alert for biases in our shortcuts, including framing. These errors can be pernicious because they occur automatically without our awareness. Errors seem most likely in certain areas. These include dealing with race, sex and gender, politics, and conflicts of interest. For example, over the past several decades, privatization and deregulation have led to more conflicted evaluation situations. When evaluators are caught in conflicts of interest, biases often distort findings. The problem is acute in the evaluation of pharmaceutical drugs. Decades ago universities conducted most drug evaluations. Now the drug companies have taken control of evaluating their own drugs. Billions of dollars depend on the findings. Such pressures distort findings, and often evaluators don’t realize they’ve made them.

Evaluators also should be prepared for racial biases. Hood and his colleagues (2015) have advanced a “culturally responsive” approach that emphasizes better understanding of minority cultures on the part of majority evaluators. Majority evaluators are often poorly informed and even misinformed about minorities. Better information can help evaluators understand cultural settings different from their own.

 Another strategy is to examine the sources of racial bias. One powerful source is the “white racial frame.” This racial frame is the way many members of the majority perceive minorities. Racial framing can bias programs, policies, and evaluations. In fact, there is strong evidence to indicate that many programs and policies have racist effects. Evaluators may overlook these racist effects and even contribute to them inadvertently. Understanding how racial framing works in detail would help us anticipate, discover, and deal with such biases.

Cognitive researcher Keith Stanovich (2011) contends that System 2 thinking should be further divided into algorithmic and reflective thinking. Algorithmic processing is what intelligence tests measure. Reflective processing is similar to what some call critical thinking, which is not measured by standardized tests. Stanovich describes reflective thinking dispositions this way:

“… the tendency to collect information before making up one’s mind, the tendency to seek various points of view before coming to a conclusion, the disposition to think extensively about a problem before responding, the tendency to calibrate the degree of strength of one’s opinion to the degrees of evidence available, the tendency to think about future consequences before taking action, the tendency to explicitly weigh pluses and minuses of situations before making a decision, and the tendency to seek nuance and avoid absolutism” (Stanovich, 2011, p. 36).

I have one other reflective disposition to add. Ordinarily, evaluators have a single framework they employ in conducting evaluations. Reflective thinking entails being able to perceive, consider, and act on two or more frameworks simultaneously and to reframe situations when advantageous. Flexibility in framing and reframing are critical aspects of rationality. Sophisticated evaluators should be able to use multiple frameworks to guide them through evaluations and be aware of the frameworks they employ. In a sense, reflectivity is a supra-evaluative set of skills and dispositions that supplements and cross checks the evaluator’s basic approach.

To summarize, when I started conducting evaluations long ago, I had no evaluation training. I had the idea that I should use the methods of the social sciences to collect data. I never thought much about why we used such methods. My current notion of evaluation is that humans are natural evaluators. Thinking is basically evaluative. Also, our thinking processes are complex and contain flaws. As professional evaluators, we evaluate naturally and employ methods to minimize biases that distort findings. Another step forward would be to become more reflective about our biases, cognitive frames, and thinking processes (House, 2015).

Hood S, Hopson R, Frierson H 2015 *Continuing the journey to reposition culture and cultural context n evaluation theory and practice* Charlotte NC: Information Age Publishing

House ER 2015 *Evaluating: Values, biases, and practical wisdom*. Charlotte NC: Information Age Publishing

Kahneman D 2011 *Thinking fast and slow* New York: Farrar Strauss Giroux

Stanovich K *Rationality and the reflective mind* New York: Oxford