Adv in Health Sci Educ (2014) 19:291–295
DOI 10.1007/s10459-014-9517-5

EDITORIAL

The Bias in researching cognitive bias

Geoff Norman

Published online: 6 June 2014

Consider, for a moment, the following two scenarios (Sherbino et al. 2012; Kahneman 2011):

(A) Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student she was deeply concerned with issues of discrimination and social justice and also participated in antinuclear demonstrations.
Which is the most likely statement?

(a) Linda is a bank teller
(b) Linda is a feminist
(c) Linda is a bank teller and active in the feminist movement

(B) Rahim is a 55 year old male who presents to the emergency department with multiple injuries following a car accident. On examination he has diminished breath sounds on the left side and a tender abdomen. His blood pressure is 90/55 and his pulse is 135 beats per minute.
Which is the most likely statement?

(a) Rahim has a tension pneumothorax
(b) Rahim has a ruptured spleen
(c) Rahim has a tension pneumothorax and a ruptured spleen

The two problems are mathematically and formally equivalent. As Daniel Kahneman, the Nobel prize winning psychologist explains in his book, Thinking Fast and Slow (2012), if the probability of being a bank teller/having a pneumothorax is p(A), and the probability of being a feminist/having a ruptured spleen is p(B), then the probability of having both is p(A) × p(B), which by definition is always less than the smaller of p(A) and p(B). Yet a majority of undergraduate psychology students will choose the normatively incorrect

G. Norman (✉)
McMaster University, Hamilton, ON, Canada
e-mail: norman@mcmaster.ca

conjunction, and therefore fall prey to the "conjunction fallacy." No one knows how many physicians will also pick the conjunction in the Rahim problem; hopefully most, so that the patient has a chance of a complete recovery.

The Linda problem is a paradigmatic example of the "Heuristics and biases" research program of Tversky and Kahneman (1974, 1984), which has identified multiple biases to which poor benighted *home sapiens*, with their inadequate mental resources, can fall prey.

Not surprisingly the program has had its critics. Perhaps one of the most erudite, yet least known, is Lopes (1991), whose criticism of the Tversky and Kahneman program is far-reaching. She points out that, as a consequence of their research, the theme of decision-making research has shifted from general support of human decision-making to scathing criticism of its inadequacies. Her critique is blunt:

> Somehow the message of irrationality has been sprung free from its factual supports, allowing it to be seen entire, unobstructed by the hopeful assumptions and tedious methodologies that brace up all laboratory research. How this happened is very complex but we can get an inkling by focusing on three related points: first, how the logic of the original experiments is ambiguous; second, how this ambiguity allows conclusions that are unwarranted to be drawn from the experiments; and third, how the unwarranted conclusions get amplified and extended by authors outside psychology. (p.67)

She seems to foreshadow the fact that 15 years later, Kahneman would win a Nobel Prize *in economics*. Her central point is that many of these demonstrations of apparent irrationality are derived from experiments that are carefully designed to create situations where commonsense would indeed fail, but then proceed to ignore these boundary conditions and proclaim the universality of human cognitive frailty.

There may be no better example of the way the "conclusions get amplified and extended" than the intrusion of these ideas into medical education. The notion that physicians, like humans, can fall prey to cognitive biases has become a mainstay of writing about clinical reasoning and diagnostic error. It is now almost axiomatic that diagnostic errors result from cognitive biases, which are firmly rooted in intuitive (System 1) thinking and corrected (hopefully) by explicit, rational, logical (System 2) thinking. These claims have been made repeatedly by numerous authors since 1999 (Elstein 1999; Klein 2005; Redelmeier 2005; Croskerry 2000, 2003, 2006).

The evidence to support these claims is surprisingly sparse. Certainly there are a number of studies (Mamede et al. 2010; Schmidt et al. 2014) showing that if you create cases where people will be vulnerable to a particular cognitive bias, clinicians will commit more errors than when it is absent. But that begs the issue of the extent to which such cognitive biases actually result in errors with "real" cases (Eva and Norman 2005; Regehr 2004). Here the evidence is more tenuous. Graber et al. (2005) reviewed a number of cases where errors occurred and found that about 68 % had cognitive biases, usually "premature closure". This amounts to concluding for the wrong diagnosis because you didn't order the right test. But it seems to me this confuses the cart and the horse; you didn't order the right test because you were thinking of the wrong diagnosis, and that may have nothing to do with cognitive bias, as McLaughlin (2014a) describes in this issue. And as Zwaan et al. (2012) has shown in an analysis of diagnostic errors, most of these derive from "mistakes" rooted in inadequate knowledge. Further, bias, like beauty, may be in the eye of the beholder; as McLaughlin (2014a) points out, we don't know how many biases would be found in detailed inspection of charts with patients who had the correct diagnosis. We do know that the name would change—now they would be called heuristics.

In view of this almost universal acceptance of cognitive bias as the root of all evil, it may be worthwhile to examine the Linda and Rahim problems in somewhat greater detail to better understand just how humans arrive at the "wrong" decision so frequently. Let us begin with Rahim, as it appears quite evident that the normatively right answer is in fact the wrong answer when you consider patient outcome. Clearly Rahim will be much better off if he is treated for both ruptured spleen and pneumothorax.

But the mathematics is unassailable. The probability of both is less than the probability of either alone. What is going on here? The answer is that a simple consideration of probability is insufficient to understand the decision-making process, which is one of Lopes' central points .

Even within a formal analytical framework, it is evident that some value—utility—should be placed on the various outcomes. For example, if we used mortality rates, we might compute the expected mortality if we diagnose (and manage) only pneumothorax, only ruptured spleen and both. Clinical colleagues estimated that, with only this information, the likelihood of a ruptured spleen is .25 and of a pneumothorax is .70. And if you treat the pneumothorax only, chance of death is .40, if you treat only the spleen, chance is .90, if you treat both mortality is .15. This enables us to examine the decisions taken under two assumptions—probability only (Kahneman's normative solution) and probability × utility (where utility is expressed as mortality and higher is worse. (The calculations for the latter are complex as you have to consider the diagnostic decision and the actual state, so will not be described here).

Decision	Probability	Prob × Mortality
D× (and R×) Spleen only	.25	.83
D× (and R×) Pneumothorax only	.70	.27
D× and R× both	.14	.16

So Kahneman is right, the likelihood of "both" is lower than either alone. But he's wrong—the chances of dying are substantially greater if you only treat one condition.

While it seems unlikely that Linda is amenable to the same logic, perhaps that is too narrow a perspective. If we accept that, while being a feminist might give her great fulfillment but is unlikely to pay the bills, and working as a bank teller achieves the opposite, and if we consider, like a true Millsian utilitiarianist, that happiness derives from both emotional fulfillment and a warm house and full belly, suddenly the conjunction fallacy is no longer a fallacy. The best life comes from being a bank teller and feminist.

Although the two problems are formally equivalent, at least in terms of the mathematics, there are some major contextual differences. For example, one needs little more than knowledge of the English language to understand, if not "solve", the Linda problem. By contrast, while Rahim is unlikely to present any diagnostic challenge to the average emergency medicine physician, he might be a very difficult problem for medical students. That is, since the Linda problem does not require specialized knowledge, errors (if they are errors) cannot be ascribed to knowledge gaps and must result from some generalizable defects like cognitive biases. Conversely, successful resolution of Rahim requires a great deal of prerequisite knowledge of clinical diagnosis, diagnostic tests and therapeutic interventions, and errors may result from knowledge gaps.

This critical dependence on specialized knowledge is omitted from most discussions of diagnostic error. Without it, we receive a distorted picture and ultimately arrive at a paradox. On the one hand, according to Kahneman, cognitive biases are universal and irremediable:

> "What can be done about biases?… How can we improve judgments and decisions…? The short answer is that little can be achieved without a considerable investment of effort. …System 1 is not readily educable…."(Thinking Fast and Slow, p.417)

On the other hand, an extensive literature shows that, as one becomes more expert and older, one relies more and more on the ostensibly error-prone intuitive reasoning of System 1 (Eva 2002). How then can it be that experts become more expert, since, with greater reliance on System 1, they will be more vulnerable to cognitive bias than novices? The answer is, of course, that experts bring more extensive knowledge, both intuitive and formal, to the problem situation. That's is why they are experts. Trying to dismiss this central part of their expertise in favour of a vulnerability to cognitive bias is a very small and flawed part of the story.

The review article by McLaughlin (2014b) in this issue expands the discussion about the role (or not) of heuristics in clinical reasoning and diagnostic errors. He makes two fundamental points, using both examples and evidence: (1) Just because a retrospective analysis of a diagnostic error reveals a likely application of a heuristic does not mean the heuristic caused the error. We d not know how many heuristics arise when the diagnosis is correct. (2) Errors may be more related to the role of different knowledge leading to good and bad outcomes. The same heuristic, based on good or bad knowledge, may lead to a good or bad outcome. Wood (2014a), in another review article in this issue, takes us out of the domain of diagnostic error and examines the role of intuitive (System 1) processes in rater judgments. In the review, he finds very clear evidence that rater judgments based on first impressions ("thin slice judgments") are usually more valid than those based on a more systematic search; moreover some of these occur with astonishing speed, <100 ms.

Wood (2014b) reappears in this issue in a commentary of an earlier paper and takes issue with the heuristics and biases literature. As we also identified above, "many of these errors were based on artificial manipulations specifically designed to invoke the actual error…" He makes a plea for a refocus away from a cataloguing of biases to "accuracy-based research" directed at "identifying factors that influence when and how people are accurate." In this respect, he harkens back to the "pre-Kahneman" era described by Lopes. Finally, we have an original study of heuristic versus Bayesian reasoning (Hall et al. 2014) and a commentary on Wood's rater judgment article.

Sometimes it appears that we, in medical education, are particularly vulnerable to urban myths that arise from minimal or inadequate evidence, and then assume a life of their own. We are likely no more vulnerable than anyone else, but it does dictate that we maintain a healthy skepticism about many apparently self-evident truths. As Hank Roediger (2013) said recently:

> "The field of education seems particularly susceptible to the allure of plausible but untested ideas and fads (especially ones that are lucrative for their inventors). One could write an interesting history of ideas based on either plausible theory or somewhat flimsy research that have come and gone over the years. And….. once an idea takes hold, it is hard to root out."

References

Croskerry, P. (2000). The cognitive imperative thinking about how we think. *Academic Emergency Medicine, 7*(11), 1223–1231.

Croskerry, P. (2003). The importance of cognitive errors in diagnosis and strategies to minimize them. *Academic Medicine, 78*(8), 775–780.

Croskerry, P. (2006). Critical thinking and decision making: Avoiding the perils of thin-slicing. *Annals of Emergency Medicine, 48*(6), 720–722.

Elstein, A. S. (1999). Heuristics and biases: Selected errors in clinical reasoning. *Academic Medicine, 74*(7), 791–794.

Eva, K. W. (2002). The aging physician: Changes in cognitive processing and their impact on medical practice. *Academic Medicine, 77*(10), S1–S6.

Eva, K. W., & Norman, G. R. (2005). Heuristics and biases—a biased perspective on clinical reasoning. *Medical Education, 39*(9), 870–872.

Graber, M. L., Franklin, N., & Gordon, R. (2005). Diagnostic error in internal medicine. *Archives of Internal Medicine, 165*(13), 1493–1499.

Hall, S., Phang, S.H., Schaefer, J.P., Ghali, W., Wright, B., McLaughlin, K. (2014). Estimation of post-test probabilities by residents: Bayesian reasoning versus heuristics? Advances in Health Sciences Education Jan 22. [Epub ahead of print].

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist, 39*(4), 341.

Klein, J. G. (2005). Five pitfalls in decisions about diagnosis and prescribing. *BMJ British Medical Journal, 330*(7494), 781.

Lopes, L. L. (1991). The rhetoric of irrationality. *Theory and Psychology, 1*(1), 65–82.

Mamede, S., van Gog, T., van den Berge, K., Rikers, R. M., van Saase, J. L., van Guldener, C., et al. (2010). Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. *JAMA, 304*(11), 1198–1203.

McLaughlin, K. (2014) Reexamining our bias against heuristics. *Advances in Health Sciences Education,* 19(3), [Epub ahead of print].

McLaughlin, K. (2014) Are we willing to change our impression of first impressions? *Advances in Health Sciences Education,* 19(3), [Epub ahead of print].

Redelmeier, D. A. (2005). The cognitive psychology of missed diagnoses. *Annals of Internal Medicine, 142*(2), 115–120.

Regehr, G. (2004). Self-reflection on the quality of decisions in health care. *Medical Education, 38*(10), 1025–1027.

Roediger, H. L. (2013). Applying cognitive psychology to education translational educational science. *Psychological Science in the Public Interest, 14*(1), 1–3.

Schmidt, H. G., Mamede, S., van den Berge, K., van Gog, T., van Saase, J. L., & Rikers, R. M. (2014). Exposure to media information about a disease can cause doctors to misdiagnose similar-looking clinical cases. *Academic Medicine, 89*(2), 285–291.

Sherbino, J., Dore, K. L., Wood, T. J., Young, M. E., Gaissmaier, W., Kreuger, S., et al. (2012). The relationship between response time and diagnostic accuracy. *Academic Medicine, 87*(6), 785–791.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131.

Wood,T. (2014a) Exploring the role of first impressoons in rater-based judgments. *Advances in Health Sciences Education,* 19(3).

Wood,T. (2014b) Is it time to move beyond errors in clinical reasoning and discuss accuracy? *Advances in Health Sciences Education,* 19(3).

Zwaan, L., Thijs, A., Wagner, C., van der Wal, G., & Timmermans, D. R. (2012). Relating faults in diagnostic reasoning with diagnostic errors and patient harm. *Academic Medicine, 87*(2), 149–156.