

Correspondence concerning this comment should be addressed to James M. Wood, Department of Psychology, University of Texas at El Paso, El Paso, TX 79968. E-mail: jawood@utep.edu

DOI: 10.1037/0003-006X.60.6.658

**On the Merits of Clinical Judgment**

Howard N. Garb

Lackland Air Force Base

William M. Grove

University of Minnesota, Twin Cities Campus

Westen and Weinberger (October 2004) criticize academic clinical psychologists for being cynical about clinical judgment and clinical practice:

[A skeptical attitude] is rapidly becoming modal among academic clinical psychologists, who, unlike [Paul] Meehl, tend to believe that they (and others) have little to learn from clinical practice or experience and whose attitudes tend to reflect what might be called clinicism (cynical toward, and negative stereotypes of, clinicians). (p. 601)

To support this criticism, Westen and Weinberger (2004) do not cite any studies or surveys, although they do offer an anecdote (p. 601).

In our view, it seems unlikely that more than a few academic clinical psychologists believe that they have little to learn from clinical practice or experience. In this comment, we examine the arguments about clinical judgment made by Westen and Weinberger (2004).

Westen and Weinberger (2004) conflate the effect of training with the effect of experience. They argue that “if one wants to know whether clinical experience confers any advantage in making broad diagnostic or prognostic judgments . . . one should compare informal predictions made by experienced clinicians with informal predictions by laypeople” (p. 607). This statement is misleading because clinicians and laypeople will differ in training as well as experience. To learn about the effect of experience, one should instead compare clinicians who have completed graduate school training and who differ in amount of clinical experience.

Westen and Weinberger (2004) do not mention that the value of training in psychology has been well-supported by research. For example, depending on the task and the type of assessment information, psychologists have been more accurate than lay judges (Aronson & Akamatsu, 1981; Garb, 1989, 1998). Similarly, psychologists with specialized training (e.g., neuropsychologists) have been more accurate for specified tasks (e.g., detecting brain impairment) than other psychologists (e.g., clinical psychologists; Garb, 1988).

The distinction between training and experience is important because the value of training has been consistently demonstrated, whereas results on clinical experience have been less positive. When psychologists and other mental health professionals are given assessment information and instructed to make particular types of ratings, less experienced clinicians generally do as well as those who are more experienced (Garb, 1989, 1998; Garb & Schramke, 1996). For example, in one study (Walters, White, & Greene, 1988), when the task was to use the Minnesota Multiphasic Personality Inventory to detect malingering in a prison setting, a group of psychology graduate students who had received training in the use of the Minnesota Multiphasic Personality Inventory were just as accurate as staff and expert psychologists. At the same time, there is general agreement that clinical experience can be helpful for generating hypotheses and ideas. Also, recent evidence suggests that clinical experience can help clinicians to structure tasks (Brammer, 2002).

Westen and Weinberger (2004) make positive comments about the types of feedback that clinicians receive: “In our clinical experience, patients routinely say things like, ‘No, I think you’re misunderstanding me,’ ‘That’s not really right,’ or ‘I don’t think what we’re doing is helping me’” (p. 603). Of course, astrologists weigh the same types of feedback when they decide whether their interpretations are correct. For a number of reasons, including the Barnum effect, psychologists can be misled by feedback. In addition, some therapists may inaccurately interpret verbal client feedback as reflecting resistance or defense mechanisms, whereas other clinicians may not recognize resistance but may instead take clients at face value.

Westen and Weinberger (2004) also argue that “psychotherapists tend to have much more direct and immediate feedback than most other medical practitioners, who may prescribe a medication or perform a procedure and not see the patient again for a year” (p. 603). But when psychologists make a diagnosis or describe a personality trait, they frequently do not receive “direct and immediate feedback” on whether they are right or wrong. In contrast, physicians often receive highly valid feedback. For example, if they believe that someone may have an HIV infection, testing can provide near-definitive feedback.

Finally, in discussing the value of ratings made by clinicians, Westen and Weinberger (2004) observe that “empirically, we have found surprisingly little evidence of theory-driven observational bias in using clinician-report methods” (p. 601). The issue is important because Westen and Weinberger argue in favor of using clinician ratings to construct diagnostic criteria. If clinicians’ ratings are biased, then the criteria will be biased. Although Westen and Weinberger report that their data indicate little evidence of bias, it is important to note that many studies on clinical judgment indicate that cognitive processes (such as remembering what symptoms a client has) are influenced by the implicit theories of clinicians (e.g., Garb, 2005; Kim & Ahn, 2002).

In conclusion, we know of no evidence that indicates that more than a few academic clinical psychologists believe that little or nothing can be learned from clinical experience and practice. Although we agree that important ideas can be generated from clinical practice and experience, we encourage clinicians to become familiar with research on the limitations and strengths of clinical judgment.

**REFERENCES**


Walters, G., White, T., & Greene, R. (1988). Use of the MMPI to identify malingering and exaggeration of psychiatric symptomatology in...

Correspondence concerning this comment should be addressed to Howard N. Garb, Wilford Hall Medical Center, Lackland Air Force Base, 59 MDOS/MMCPR, 2200 Bergquist Drive, Suite 1, San Antonio, TX 78236-5300. E-mail: howard.garb@lackland.af.mil

DOI: 10.1037/0003-066X.60.6.659

Clinical Judgment in Science

Drew Westen
Emory University

Joel Weinberger
Adelphi University

We appreciate the three comments on our original article (Westen & Weinberger, 2004) and our opportunity to reply to them. Schulz and Waldinger (2005) make the important point that just as researchers can capitalize on the knowledge of experienced clinical observers through aggregation, they can aggregate the judgment of lay observers in assessing phenomena such as emotion. The reason, as they articulate, is that skills such as “reading” emotion from facial expression, tone of voice, posture, and the constellation of cues provided in everyday life are an area of expertise for most people, one that is now often called social or emotional intelligence. As psychometricians have known for years, one can increase reliability in many different ways (e.g., by refining a coding scheme, increasing the number of items to be rated, or increasing the number of raters). Block (1971), Ambady and Rosenthal (1992), and others have demonstrated the benefits of harnessing the observational powers of untrained young adult observers, whose aggregated ratings can provide data that are both reliable and valid in many domains. Clearly, a task for future research is to identify those circumstances in which aggregated lay reports are as useful as or more useful than coding manuals or aggregated data from other observers.

The comments by Wood and Nezworski (2005) and Garb and Grove (2005) do not address our central thesis—namely, the importance of distinguishing two meanings of clinical: The first, appropriately attributed to Paul Meehl, refers to a mode of informal, subjective judgment; the second refers simply to the mental processes and products of clinicians (experienced practitioners). We argued that failure to distinguish these meanings has led to an unfortunate tendency to view clinicians as somehow uniquely defective in their thinking. Reviewing a large body of empirical data, we showed that clinicians can in fact make highly reliable and valid judgments under circumstances that maximize their expertise (e.g., using psychometric instruments, making judgments at an appropriate level of generalizability), and we argued that future work might gain from focusing less on documenting the failures of clinicians and more on identifying the circumstances that are likely to maximize the appropriate uses of their skills.

We presume that Wood, Nezworski, Garb, and Grove do not find our primary thesis objectionable, given that they do not mention it. From reading Wood and Nezworski’s (2005) comment, one might think that our article was a philosophical treatise on the futility of science. They characterize it as arguing that the history of science is the history of confirmatory biases and contrast this with their more hopeful view of science. This seems to us an odd rendering of our article, given that both of us are scientists who have spent much of our professional lives collecting data (and presumably would not have done so if we believed such efforts to be futile), and our article was a review of scientific data, not of philosophical argument.

The point of the sentence around which Wood and Nezworski (2005) build their comment was simply that the same biases widely attributed to clinicians are common in scientists as well—a point for which we would be delighted to take credit, but it is one that was actually made much more elegantly by the historian and philosopher of science Thomas Kuhn (1962). Kuhn’s classic work is filled with examples of such biases. We suspect that we share with Wood, Nezworski, Garb, and Grove a faith that, in the long run, such biases tend to be overcome. But as John Maynard Keynes (1924) once said, in response to those who argued that economic cycles average out in the long run, “in the long run we are all dead” (p. 88). Gregor Mendel might take solace in knowing that his extraordinary experiments, which formed the foundation of modern genetics, were recognized for their importance years after his death, but only if his faith in both science and an afterlife were confirmed. It behooves us, especially as psychologists, to be on guard for such biases in science. We argued that the only way to lessen such biases, to which all of us are prone, is through self-reflection and knowledge of the tricks our minds can play on us—a point with which we assume we all agree.

After taking out of context a single sentence from a lengthy review of scientific evidence, Wood and Nezworski (2005) complain that we misrepresented their work and admonish us that the “science of psychology will progress most quickly if disagreements are conducted in a respectful manner, with close attention to accuracy” (p. 657). We could not agree with that statement more. In that spirit, we would respectfully suggest that they reread some of their own published critiques of the Rorschach, which liberally mix worthy arguments, one-sided presentations of evidence, and innuendo about the scientific integrity of researchers who disagree with them (cf. Meyer, 2000). Instead of engaging in blanket faultfinding, perhaps we can all agree to be respectful of the possibility of genuinely different readings of the available evidence (allied with recognition of the usual amount of motivated reasoning that accompanies the study of virtually anything about which scientists, like other people, are passionate). The role of much of scientific investigation is in fact to design the next study in the hope of resolving just such differences of scientific opinion.

We will, however, respond to Wood and Nezworski’s (2005) specific concerns about misrepresentation, because we certainly did not intend to misrepresent their or anyone else’s work, and we agree that scholarly arguments are best conducted without caricature of alternative viewpoints. Wood and Nezworski (2005) take issue with our characterization of their comments on a meta-analysis by Hiller, Rosenthal, Bornstein, Berry, and Brunell-Neuleib (1999) comparing the relative effect sizes of Rorschach and MMPI indices (Garb, Wood, Nezworski, Grove, & Stejskal, 2001) as follows: “Claiming that confirmatory bias characterizes science, Westen and Weinberger (2004) criticize our commentary on this meta-analysis” (p. 657). Our argument was actually not that scientists are incapable of transcending such biases but, as stated above, that they need to practice systematic self-reflection to minimize the impact of such biases. We suggested that Garb and colleagues dismissed the Hiller et al. meta-analysis with a series of post hoc methodological complaints that, taken together, suggested that Hiller et al. did not understand the basics of meta-analysis. Aside from asserting that Hiller et al. did not correctly calculate reliability and appropriately address the issue of blind ratings, Garb et al. (2001) also admonished them for reporting unweighted along with weighted effect size estimates—suggesting that doing so is not “proper” (p.