

Out, damned blot!

By Frederick Crews

Review of: *What's Wrong with the Rorschach?: Science Confronts the Controversial Inkblot Test*. By James M Wood, M Teresa Nezworski, Scott O Lilienfeld and Howard N Garb. 446pp. Jossey-Bass. \$44.95.

From *The New York Review of Books*. ©2004.

It's a Rorschach. That bit of everyday speech, referring to any equivocal stimulus that elicits self-betraying interpretations on all sides, is one sign among many that, in the popular mind at least, the vaunted inkblot challenge has no rival as psychology's master test. In actuality, the Rorschach is now administered for diagnostic purposes somewhat less frequently than the low-maintenance, question-and-answer Minnesota Multiphasic Personality Inventory, which asks the subject to agree or disagree with such flatfooted assertions as "I often feel sad". But neither the public nor Rorschachers, as the zealous and clannish guardians of the blot technique are known, take much interest in "superficial" self-report tests such as the MMPI. The mind's hidden layers, it is assumed, can be tapped only through unguided responses to images lacking determinate content; and the Swiss psychiatrist Hermann Rorschach's 10 cards with bisymmetrical shapes, introduced to an initially unimpressed world in 1921, are thought to have confirmed their uncanny power in countless applications.

This judgment is shared in large measure by American clinical psychologists and other professionals who have occasion to administer personality tests. As we learn from a provocative and important book by James M Wood, M Teresa Nezworski, Scott O Lilienfeld and Howard N Garb, *What's Wrong with the Rorschach?*¹

Some 80 per cent of American PhD programs in clinical psychology still emphasise the Rorschach in required courses; 68 per cent of specialist programs in educational psychology teach Rorschach technique; and the test is employed by roughly a third of all US psychologists evaluating parents in custody cases, criminals facing sentencing or parole, and children who may or may not have been abused. Until very recently, testimony by Rorschach experts has gone largely unchallenged in the courts.

Necessarily, then, *What's Wrong with the Rorschach?* is not just a history of the test's evolution and

periodic vicissitudes. It is also, and with increasing social concern as the story approaches the present, a continual assessment of the merits and pitfalls of projective testing. Since the four authors have themselves been participants in recent debate about the Rorschach, there is no pretense of neutrality here. But Wood and his colleagues do aim at objectivity and fairness, and if they err at all, it is on the side of mercy. Readers of *What's Wrong* will find no more lucid primer on the requirements of scientific prudence as they relate to the authentication of psychological tests.

An avid reader of both Freud and his Zurich colleague Jung, Rorschach conceived of his test as a non-sectarian aid to psychoanalysis, impersonally determining an individual's "experience type" (*Erlebnistypus*) without presuming to favour one psychodynamic faction or another. The idea was to present all test takers with the 10 loose printed cards, half of them in black and white and half including some colours, displaying an identical sequence of images.²

For each card the test giver would ask in the most neutral tone, "What might this be?"; he would capture the subject's responses fully and exactly in a "protocol", or written record; and he would subsequently arrive at a singular personality profile by sifting that record for telltale features such as the kinds of forms named, whether the focus was on whole shapes or details, emphasis on movement versus colour, and whether there was a rigid literalism or a comfortably imaginative accommodation to the imperfect resemblance between the blots and any real-life form.

The extent to which responses emphasized movement and color was of paramount importance in Rorschach's system of weighing personality. He believed that test takers who offer a high number of movement ("M") responses are, paradoxically, turned inward or "introversive"; intelligent and creative, they nonetheless are awkward and socially inept. In contrast, sub-

jects who favor color ("C") responses are "extratensive", or adroit in company but restless and impulsive. Someone who registers a high number of both M and C scores qualifies as "diluted" or "ambiequal" - a healthy blending of introversive and extratensive traits. But low and similar numbers of M and C responses stigmatise the subject as "coarctative", or lacking in both creativity and emotional stability.

These rules were far from modest in scope. The close association between creativity and social clumsiness, were it to be upheld by evidence from other sources, would in itself constitute a major discovery, and so would the posited link between social adeptness and impulsiveness. In addition, it would be remarkable if those and other constellations could be inferred with certainty from such utterances as "The bug is bleeding" and "It looks like a skull". And this is to say nothing of Rorschach's most expansive boast, which was that his test would be found capable of ascertaining personality differences between regional populations and even whole races.³

Did he have grounds for making such sweeping claims, or was he capriciously assigning the equivalent of fortune cookies to his unsuspecting volunteers? There is much in Rorschach's only book, *Psychodiagnostics*, that might encourage us to regard him as a crank. Bizarrely, for example, he insisted that a movement response be scored if the subject conjured a child sitting at a desk or a vampire sleeping in a coffin, because "muscular tension" was supposedly implied. And although a dog performing in a circus exhibited Rorschach movement, a cat catching a mouse or a fish darting through water did not, because, according to the founder, significant motion had to be "human-like" in function. Meanwhile, Rorschach tagged as "pedants" or "grumblers" any test takers who concentrated on details as opposed to whole images; those who interpreted white spaces were probably troublemakers; and those who hesitated before commenting on the multicolored cards must be exhibiting "color shock", thereby betraying themselves as neurotic repressers of emotion.

Rorschach argued, quite sensibly, that by examining the average test results (later called "norms") for many people whose personality traits had already been determined by other means, administrators could learn whether a given kind of response was actually well correlated with a given trait. Yet before he died from a perforated appendix just nine months after the publication of *Psychodiagnostics*, Rorschach had found time to accumulate test results for only 405 independently categorized subjects, and their types were drastically skewed towards schizophrenia (188 examples) and other pa-

thologies he had encountered in his hospital rounds. Only 117 ordinary people, scattered amid his assorted "morons", "imbeciles", "senile dements" and so forth, had been sampled. Given such a sketchy evidential base, it isn't surprising that *Psychodiagnostics* was slow to find admirers; the wonder is that its complex scoring system was finally adopted with so few reservations.

The Rorschach found its true welcome in the world's headquarters of psychological typecasting and "adjustment", the US. Wood engagingly tells how the test finally caught on there in the 1930s, flourished in the 1940s and early 1950s, weathered a crisis of doubt in the later 1950s and 1960s, and then surged again until, beginning a decade ago, skeptics began to nip at its heels once more. Along the way, different groups of US enthusiasts devised their own scoring rules to yield the kinds of results that interested them. Through it all, however, Rorschachers have kept faith with the founder's 10 inviolate cards, which have been granted the kind of awe once reserved for texts dictated directly from the sky.

The Rorschach conquered North America and much of the Western world before any part of its rationale had been subjected to stern experimental trial. In seeking to explain this striking fact, Wood notes that inkblot games and tests were current even before Rorschach launched his own version. His key departure - the attempt to gauge a subject's whole personality and not just a faculty of imagination - fit nicely with the growing sway of psychoanalysis, and particularly with the Freudian ideas of projection and free association. Again, those who preferred the more cheerful Jungian conception of the psyche responded favorably to Rorschach's adaptation (with significant differences) of Jung's already celebrated dichotomy between introverts and extroverts. Moreover, in balancing a "romantic" emphasis on deep intuition against an "empiricist" battery of codes and tables for figuring scores, the Rorschach proved at first serviceable, and then virtually indispensable, to the burgeoning American profession of clinical psychology, which was developing its own romantic pretensions but needed an objective-looking diagnostic tool to offset the inherent subjectivism of the one-on-one interview.⁴ Reasons for popularity, of course, are not the same thing as scientific justifications. Wood *et al* remind us that if a given instrument of testing in any field is not to cause havoc, it must be both valid and reliable.

In brief, it must measure what it purports to measure and it must yield approximately the same results when readministered in new conditions or by other examiners.

Rorschach had accepted those criteria in principle, and most of his followers have paid due obeisance to them. But at every juncture where the test stood in peril of being decertified by negative findings, Wood shows, its promoters backed off from empirical accountability and expanded the scope of their claims. The story told in *What's Wrong*, by turns appalling and amusing, reads like a parable of the larger struggle between science and pseudoscience, with the latter always managing somehow to issue itself a new reprieve from execution.

When the Rorschach began to attract American followers through word of mouth in the 1930s, it brought to prominence an initially reluctant but subsequently flamboyant champion, Bruno Klopfer, whose talent for salesmanship and deafness to criticism were responsible in part for the high morale of American Rorschachers in the 1940s and 1950s. A refugee from Nazi Berlin, Klopfer had studied with Jung in Zurich and had learned how to score various psychological tests there, including the Rorschach. He was barely surviving as a research assistant in Columbia's anthropology department when eager graduate students learned of his expertise and pressed him into moonlighting as their Rorschach trainer.

Although Klopfer's real passion had been psychoanalysis, not assessment, he soon contracted a taste for interpreting Rorschach's still untranslated pronouncements and for devising novel inkblot rules that hadn't occurred to the master. Before long he possessed a grand career and an adoring crew of disciples who fed his insatiable ego. As Wood explains, this elevation of one person to guruhood added mystification to an already dubious mind-reading program and further postponed a reckoning with the need for evidential support.

In the Rorschach scheme as first conceived, the sum of a subject's scores for responses in each category of interest - color, say, or white-space shapes - corresponded directly to a certain trait of personality. Klopfer accepted some of those equivalences, but on the whole he found the idea behind them too rigid for capturing the subtleties of human character. What was needed, he argued, was "configural" interpretation, whereby a highly experienced and gifted judge (guess who?) would draw "holistic" inferences from an intuitive contemplation of all of the subject's scores on the test. The Rorschach judge or "artist" could justify this method by creating anonymous ("blind") profiles on the basis of protocols compiled by others and then by checking the profiles against case histories or against delayed personal acquaintance with the test takers. The artist himself or someone from his circle of admirers would let the rest of us know, anecdotally, how well he had done.

Klopfer's fans considered him a virtual oracle, and he was inclined to agree. He even claimed that, through analysis of Rorschach scores alone, he could discriminate between cancer patients with fast- and slow-growing tumors. But all of his pretensions were proved hollow when neither he nor other famous virtuosos could exhibit any diagnostic acumen in circumstances that were properly secured against cheating. Though many Rorschachers still revere Klopfer's memory, he now appears to have been only a colourful buffoon.

Wood and his fellow research psychologists are only peripherally concerned with Klopfer's foibles. Their target here is the whole idea of "clinical validation", whereby hypotheses are checked not against impersonal trials of their adequacy but against testimonials, case studies and assessments of success made by parties with a stake in the outcome. That method risks being undermined by "confirmation bias", or the natural human tendency to misread evidence in one's favour. And confirmation bias runs wild in subjective evaluations of Rorschach profiles, thanks to such factors as the test's excessively broad classification of traits and the contradictions that crop up within a given subject's scores, tempting the evaluator to seize upon apparent hits and ignore the misses. Even blind Rorschach interpretation *per se*, Wood points out, isn't always what it seems, because the examiner often has advance knowledge about the population of test takers - for example, a ward of mental patients.

Klopfer's disdain for controlled studies was shared by American psychoanalysts, whose own absolute trust in clinical validation had come straight from Freud. The Rorschach, they perceived, could serve as a technical adjunct to their relatively unstructured explorations of patients' minds. Predictably, they imbued the test with a symbol-decoding function - an approach that Rorschach himself, decades earlier, had pondered and rejected as unproductive. According to the Freudians, subjects gazing at the 10 plates were really seeing projections of their unconscious desires and neuroses - conditions that pointed to a need for further months or years on the couch. And needless to say, those inferences, too, were clinically validated without incurring any risk of disconfirmation.

At the height of the Freudian vogue, Rorschach's Cards IV and VII became known to analytically inclined authorities, though not to unsuspecting test takers, as the "Father" and "Mother" images.⁵ A woman who pointed out, plausibly, that the "arms" on the Father card are skinny was said to be in the grip of penis envy; and if she likened the Mother to a stuffed animal, she thereby convicted herself of what one expert called "a refusal to grow up and assume heterosexual responsi-

bilities". "So it went with all the cards," Wood observes. "Card V revealed childhood memories of having seen one's parents engaged in intercourse. Card VI reflected unconscious attitudes toward sex and 'phallic worship.' Card IX revealed 'anal' concerns and paranoia. Card X revealed 'oral' fantasies."

In this vein of instant diagnosis no one surpassed the psychoanalyst Robert Lindner, the author of *Rebel Without a Cause*. Lindner identified 43 Rorschach responses amounting to cries for help. Does the subject perceive "some sort of tool" in Card II? Then he is suffering from "hesitancy in coming to grips with an underlying sexual problem". Does he compare a portion of Card X to an extracted tooth? He is a chronic masturbator. And if, when perusing Card IV, he is rash enough to mention both decay and death, he is probably suicidal, and "there is a fair prospect that [he] will benefit from convulsive therapy".

All of the eminent US Rorschachers from the 1930s through the 1950s were sympathetic to Freudian dream interpretation, and they felt the tug of the popular psychoanalytic tide. After considerable hesitation, for example, and not without misgivings about diluting his authority, Klopfer made room in his scoring system for Mother and Father symbolism. Yet the main benefit of the Freudian trend accrued to those tradition-minded Rorschachers who resisted it. In doing so, they identified themselves as the party of scientific restraint even though their own rules for interpreting color, movement and form responses as indicators of personality had never been proved cogent, either.

From the outset of the Rorschach's wild American ride, some psychologists who believed in the test's general soundness understood that empirical standards couldn't be indefinitely brushed aside in the lordly Klopfer manner. Two highly regarded Rorschach theorists, Samuel Beck and Marguerite Hertz, argued vigorously that Klopfer's "configural" blending of scores was perpetuating a deadly subjectivism and placing each individual rule of interpretation beyond the reach of disproof. Beck and Hertz won an appreciative following by associating themselves with "psychometrics", the statistically based controls that are now universally honoured in experimental psychology if not in its clinical counterpart. As they emphasized, the psychometric ethos mandates that test procedures be standardized; that reliability be verified, not just promised; and that norms be gathered in enough volume to put the announced meaning of scores beyond dispute.

Beck and Hertz could endorse psychometrics because they were initially sure that research would confirm most Rorschach rules while weeding out a few unsus-

portable ones. That confidence inspired many loyalists as well as outsiders, from the 1940s through the 1960s, to submit Rorschach hypotheses to objective review. But the results proved devastatingly negative. Correlations between predicted traits and independently observed ones were found to be either nonexistent or too weak to be trusted. Test results varied unacceptably among examiners with differing styles of self-presentation. And meanwhile, crippling statistical blunders were unearthed in the more optimistic reports. As the most sophisticated and persistent of the critics, Lee J Cronbach, wrote in 1956, "It is not demonstrated that the test is precise enough or invariant enough for clinical decisions. The test has repeatedly failed as a predictor of practical criteria ... There is nothing in the literature to encourage reliance on Rorschach interpretations."

Here was one of those moments when Rorschachers faced a clear if painful choice between loyalty to science and loyalty to beliefs in which they had invested much money, time and self-esteem. As usual, most of them opted to close ranks. Even Beck ignored the ominous findings and began taking a kinder view of clinical validation after all. And most of those who did refer to the experimental literature, having sifted it for scraps of encouragement, refused to acknowledge what it was plainly saying about the test's fundamental inaccuracy.

When practitioners of a quasi-medical fad exempt themselves from answerability to empirical trials, one usual consequence is a proliferation of alternative schools. By the late 1950s, Wood reports, no fewer than five American Rorschach regimens were current, not to mention other inkblot tests that heretically departed from the original cards. Credentialed psychologists, rapidly increasing in number but not in methodological sophistication, felt free to draw upon all five incompatible codes as if they were passing down the line at a salad bar. To the increasingly restive dissenters, what had been true all along was now overwhelmingly apparent: the Rorschach was a revealing projective test not of its respondents' quirks but of the preconceptions held by its advocates.

By the mid-1960s the Rorschach had earned the distrust of most psychologists who were keeping up with the mainstream journals.

In 1974, however, the technique underwent a surprising resuscitation at the hands of the American clinician John E Exner, whose often revised and supplemented book, *The Rorschach: A Comprehensive System*, would become the most influential of all Rorschach texts.

Seeking consensus among the quarrelling Rorschach-

ers, Exner assembled an eclectic quilt of the best elements, as he judged from painstaking surveys, in all of the competing regimens while adding many categories of his own, including new measures for egocentricity, depression, obsessive style and "hypervigilance". In itself, such a combination of winnowing and amplifying wouldn't have set Exner apart from many another Rorschach pundit. But he silenced most doubters by conspicuously embracing psychometric standards, including the provision of abundant, broadly representative norms and barrages of scientific references purporting to document both the validity and the reliability of his Comprehensive System.

In Exner's hands the test regained a level of respect not enjoyed since the early 1950s, and his domination of the Rorschach scene has been all but total for three decades now. In 1997 he received an "Award for Distinguished Professional Contributions to Knowledge" from the Board of Professional Affairs of the American Psychological Association, which credited his Comprehensive System with having revived "perhaps the most powerful psychometric instrument ever envisioned"; and as recently as April 2004 his life's work was honoured through a conference jointly sponsored by Harvard Medical School and the Massachusetts Mental Health Centre.

As *What's Wrong with the Rorschach?* demonstrates, however, the imposing Comprehensive System is really a production of smoke and mirrors. Exner's claims for the high reliability of his technique, it turns out, have rested on a misconstrual of accepted statistical terminology. Further, his famous compilation of norms was vitiated by a major sampling error that went unremarked for more than a decade. The cited studies underpinning his rules have been mostly unpublished and unshared work by the least trustworthy of judges, a team of enthusiasts employed by his own subsidiary, Rorschach Workshops, Inc. And the inflated reputation of those studies has been sustained by the Rorschach Research Council, yet another Exner satellite.

In Exner's code, "reflection" answers, such as "clouds reflected in a pond", indicate narcissism. Even one reflection response within a protocol, Exner admonishes, can tell a psychologist that "a nuclear element in the subject's self-image is a narcissistic-like feature that includes a marked tendency to overvalue personal worth". This is a grave matter, because Exner, taking a page from Freud, has held that "homosexuals and sociopaths" are highly narcissistic. Detection of those dubiously bracketed personality types thus rests in part on the perception of symmetries in forms that simply are symmetrical.

Wood's critique leaves me convinced that Exner's pseudo-precise method for inferring diagnoses from weighted combinations of Rorschach scores is even riskier than Klopfer's impressionistic holism. Take, for instance, the Comprehensive System's egocentricity index, which is figured by tripling the number of a subject's reflection responses, adding the number of "pair" responses, and dividing the sum by the total number of all responses.

If any one of the component rules here is invalid, so is the whole index, in which case its automatically applied mathematical formula will cause many more misdiagnoses than Klopfer's case-by-case guesswork.

The Comprehensive System has done less than nothing to remove the most serious flaw in all previous Rorschach schemes, including Rorschach's own: their tendency to overpathologize, or to err on the side of abnormal characterizations.

In one study from the 1980s, in which mental patients and ordinary citizens were blindly intermingled, Comprehensive System judges classified nearly 80 per cent of the normal subjects as depressed and as harbouring serious problems of character. And in 2000 three sceptical researchers, examining the Rorschach scores of 100 behaviourally normal California schoolchildren, reported that, according to the Comprehensive System, the children "may be described as grossly misperceiving and misinterpreting their surroundings and having unconventional ideation and significant cognitive impairment. Their distortion of reality and faulty reasoning approach psychosis.

These children would also likely be described as having significant problems establishing and maintaining interpersonal relationships and coping within a social context. They apparently suffer from an affective disorder that includes many of the markers found in clinical depression".

When the Exner system is relied upon to assess individual children or candidates for the priesthood or pilots suspended for drunkenness or convicts seeking parole, a form of roulette is being played with their fate. Wood offers a chilling example from a custody dispute in which he himself was consulted too late to affect the outcome. Inkblot scores had suggested that the ex-wife, who had repeatedly charged her ex-husband with physically and sexually abusing his children, was seriously disturbed, lacking in empathy and incapable of forming rational judgements. Those were the conclusions implied by such damning symptoms as her having seen, in two cards, the shape of a paper snowflake (incorrectly scored by the examiner as a "reflection response") and

the supposedly depressive image of a Thanksgiving turkey carcass - corresponding, as it happened, to leftovers that were then sitting in her refrigerator.

Meanwhile, the ex-husband's Comprehensive System protocol assured the authorities that he was more or less normal. As Wood learned from available records, however, the man had beaten at least one of his three previous wives, had married this fourth one under an assumed name, had broken two of her teeth shortly after their wedding, and had both battered and molested his young son. Thanks in large part to the two errant Rorschach profiles, full custody of the endangered son was awarded to his sadistic father.

If a psychological test cannot discriminate reliably between signs of pathology and casual associations such as the remembered turkey carcass, it is a public menace and ought to be dropped forthwith. Oddly, however, Wood *et al* are reluctant to say so. Partly out of deference to colleagues who have devoted their careers to the Rorschach and partly because some of the authors themselves still harbour confessed "romantic" sympathies, they take a firm stand only about the urgency of getting the Rorschach out of the courtroom. On other points, such as the possible usefulness of inkblots in "psychodynamic exploration" that could be "analogous to dream interpretation", they express cautious interest and call for further research.

Here Wood *et al* appear to have put in abeyance their own decisive critique of clinical validation. Both dreams and Rorschach responses can be "explored" with disastrous effect; think of the role played by dream analysis in recovered memory therapy, and think of Robert Lindner's suggestion that shock treatment may be indicated when Rorschach answers reveal a desperate mental state. The story told with admirable patience and logic in *What's Wrong with the Rorschach?* speaks more clearly than its authors do here at the end. This test is a ludicrous but still dangerous relic of the previous century's histrionic love affair with "depth", and the only useful purpose it can serve now is as a caution against related follies.

Frederick Crews is the principal author of *The Memory Wars: Freud's Legacy in Dispute* and the editor of *Unauthorised Freud: Doubters Confront a Legend*. His most recent book is the satire *Postmodern Pooh*.

NOTES

1 James M Wood and M Teresa Nezworski are associate professors of psychology at the University of Texas at El Paso and Dallas, respectively; Scott O Lilienfeld is associate professor of psychology at

Emory University; and Howard N Garb, formerly clinical associate professor of psychiatry in the University of Pittsburgh's School of Medicine, is now at Wilford Hall Medical Centre, Lackland Air Force Base. Since all but one of the book's 12 chapters were drafted by Wood, I will usually write 'Wood' when designating the authors collectively. No slight to the other collaborators is intended.

2 It was long thought that Rorschach created his cards simply by folding each wet, multi-blotted page in half along its vertical axis. It now appears, however, that he either painted entire shapes or subtly altered his blots with watercolours to produce desired effects.

3 Hermann Rorschach, *Psychodiagnostics: A Diagnostic Test Based on Perception*, translated by Paul Lemkau and Bernard Kronenberg (Grune and Stratton, 1975), pp. 96-97, 102, 107, 112. With the assistance of prominent American Rorschachers, a Columbia PhD candidate followed Rorschach's lead in 1939, using test results to suggest that 'the White race', sampled in one of its typical habitats, Columbia Teachers College, is more introversive on the whole than 'the Negro race', sampled in Harlem. See Mary Hunter Sicha, *A Study of the Rorschach 'Erlebniss-Typus' [sic] of Comparable White and Negro Subjects* (doctoral dissertation, faculty of philosophy, Columbia University, 1939), pp40, 56.

4 The struggle between romantic and empiricist tendencies in modern psychology is highlighted in Paul R McHugh, 'Psychotherapy Awry', *The American Scholar*, Volume 63 (1994), pp1730.

5 Respect for colleagues has impelled Wood *et al* to omit images of the cards, whose (already debatable) effectiveness could be jeopardised if they were made too accessible. But several other books do reproduce the cards, and anyone with internet access can easily find depictions of them.