The Ethical Implications of Paul Meehl’s Work on Comparing Clinical Versus Actuarial Prediction Methods

Robyn M. Dawes
Carnegie Mellon University

Paul E. Meehl’s work comparing statistical versus actuarial prediction—and the large body of research that followed by others on the same topic—was mainly theoretical and empirical. Meehl also suggested that this work led to a “practical” conclusion, which was quite strong. The author argues that, in addition, it leads to an ethical conclusion, equally strong. Whether the scientific findings are combined with an overarching ethical principle that the best predictions possible should be made for clients, or whether these findings are framed as delineating what can be done for clients—and that clinicians ought not to attempt to do what they cannot—the conclusion is the same. Whenever statistical prediction rules (SPR’s) are available for making a relevant prediction, they should be used in preference to intuition. Any modification of these rules should be systematic and subject to the same type of evaluation originally used to assess the SPR’s themselves. It is even possible to develop near-optimal rules in new situations. Providing service that assumes that clinicians “can do better” simply based on self-confidence or plausibility in the absence of evidence that they can actually do so is simply unethical. © 2005 Wiley Periodicals, Inc. J Clin Psychol 61: 1245–1255, 2005.

Keywords: Paul E. Meehl; clinical prediction methods; actuarial prediction methods; statistical prediction rules; ethical practice

Lake Webster in Massachusetts (near the Connecticut border) had a Native American name until roughly 50 years ago. That name was Chargoggagoggmancharevogog-gogcharbunagungamaug. The English translation of this rather long name is, “I fish on my side, you fish on your side, and no one fishes in the middle.” It was apparently given that
name as part of a settlement of a tribal fishing dispute involving which group could fish where.

When I entered the field of clinical psychology in 1958, the name of that lake described the field quite well. When not doing “their own thing”—as a result of their experience-based insight about people and pathologies—clinicians often adhered to a particular “school” of analysis and practice. (The analogy is to the people fishing in the lake, not to the schools of fish.) With the exception of a few occasionally intolerant and public disagreements, most practitioners adopted a laissez-faire attitude toward others, even those with different theoretical orientations. While Freudians could not quite believe that behavioral approaches would do as well as psychoanalytic ones (and vice versa), clinicians felt little need to prescribe how those with different orientations should behave. (Even sexual contact in therapy was not officially prohibited until almost two decades later.) There was conflict between different modes of treatment, but it was that of a horse race. Perhaps, after understanding “repression” well enough to be able to study it in detail and manipulate it, Freudian approaches would turn out to be the best; perhaps, after the translation from the Skinner box to the office had been perfected, behavioral therapy would win; perhaps Egon Brunswick would turn out to be correct, and the best approach would be to develop “ecologically valid” models where the predictive validity of influences, and their relationships, mirrored those found in the natural social–psychological world. Or perhaps neuropsychology would advance to the point that specific problems (or talents) could be identified with either brain structure or function, or both. (For example, drug treatment for schizophrenia was beginning to show promise at exactly the time I entered the field.) Even psychosurgery might turn out to have some sort of benefit that had previously eluded those of us who were opposed to its practice.

The “gold standard” of randomized clinical trials for evaluating effectiveness (and even now we diddle around about the differences between “effectiveness,” “efficacy,” and “efficiency”) had not been well publicized until Jonas Salk’s polio vaccine trials of 1954, and adoption studies indicating the importance of genes in understanding such disorders as schizophrenia and alcoholism had yet to be conducted. Except for the “schools of thought,” psychologists tended to be very creative—if for no other reason than that so few of them (us) tended to do the same thing. A particularly appealing characteristic of the creative laissez-faire approach was that it permitted so many of us to have a lot of fun.¹

I began worrying about the inferences of my mentors that symptoms or responses “typical” of underlying conditions could be used in an associative way to diagnosis these conditions. Because there were often so many more “normal” patients than schizophrenics in many samples, I argued that someone who gave a “typical” schizophrenic response was more likely to be an “atypical normal” than a “typical schizophrenic.” I discussed my concerns with a close friend (Professor James L. Hedegard, recently retired), and he suggested that I write an essay about the problem. I followed his advice, and even included a little mathematics together with the verbal statement. When I showed the essay to my mentors, one of them said that it was “something like this guy Meehl” did. So I was sent to the late Professor Warren Norman, who had been a student of Meehl’s at Minnesota.

¹My favorite example of having fun was first to present a very depressed client with the blank card of the TAT and ask him what was going on, and then present the same blank card again to ask what couldn’t be happening. Because the ordinary instructions elicited a family scene where everyone was happy and loving, whereas my “negation instructions” elicited a scene where family members were knifing and shooting each other, I became half-convinced that I had obtained evidence for the Freudian idea that depression was hostility turned inward, hostility that could be “released” under instructions to be unrealistic. Fortunately, it also occurred to me that because so few family get-togethers involve murderous mayhem, this client could just be “following the base rates” in constructing such a scene when the usual instructions of “what could be happening” were reversed.
After very politely and hesitantly checking out to make sure that I had not plagiarized Meehl, Warren suggested that I read a lot of Meehl’s papers and his 1954 book (Meehl, 1954/1996). Why were these not part of my curriculum already? The reason apparently was that while Norman’s course on personality assessment was held in high regard in the psychology department at Michigan, it was not required for any of the clinical students, who were instead introduced to individual differences and later psychotherapy through a course in the second year entitled “Advanced Rorschach Techniques.” In fact, I know of only one clinical student who took Norman’s course—given it was clear that Norman was not part of the “clinical group.” Another professor excluded from the clinical group was the department head, E. Lowell Kelly, despite his critical role in starting the field of clinical psychology by convincing the Veterans Administration at the end of World War II that psychologists should at least supplement if not replace psychiatrists in the Veterans Hospital system. (Kelly pointed out the unique background of psychologists due to their [our] alleged devotion to the “Boulder model” of combining research with practice; his recommendation was buttressed by the large number of returning veterans in need of some sort of psychiatric service or counseling, and the fact that at the time psychologists were a lot cheaper than were psychiatrists.)

Again, we were all fishing in different parts of the lake. The clinical people were teaching Rorschach and psychoanalytic interpretations, Norman was teaching statistically based personality assessment, Kelly was attempting to integrate research with practice by demonstrating that unstructured interviews had virtually no predictive validity (Kelly, 1953), and so on.

Particularly after reading Meehl’s 1954 book (Meehl, 1954/1996), I asked myself why we were not following its obvious implications. Moreover, I asked myself why we were not engaged in the simple two-step logic of inferring the probability of a condition, given a symptom from the probability of the symptom given the condition, combined with estimates of the relevant base rates of condition and symptom—rather than identifying (irrationally) the probability of the condition given the symptom with the probability of the symptom given the condition. I was told that this was “Meehl’s stuff,” and that although Paul Meehl was a very brilliant man—perhaps the most brilliant around—“what he does has nothing to do with what we do.” (I believe that is an exact quote.)

The lake analogy applied not only to actual practice, but to the “ways of thinking” about it as well. In a 1994 article entitled, “Uncommon Sense: The Heretical Nature of Science,” Allan Cromer writes, “Scientific thinking, which is analytic and objective, goes against the grain of traditional human thinking, which is associative and subjective. Far from being a natural part of human development, science arose from unique historical factors” (p. 688). What “we” did was regarded as a natural extension of traditional thinking. We were allegedly selected, in part, on the basis of being understanding, empathetic human beings (I don’t know how I made it), and our clinical training as projective test analysts and apprentice therapists was meant to “hone” our natural skills through experience with clinical populations. In contrast, Meehl’s work began with statistical analysis and applied this analysis to empirical results—involving groups of people. It was “unnatural” in two ways.

First, it involved generalization from groups to individuals. Most of us expected such inference in the medical field (surely we would prefer a treatment that had been proven “twice as effective” in a large group of people), but the idea of how to understand an individual was allegedly “idiographic.” There were of course statements about entire groups—such as “schizophrenics” or “depressives”—but the conception was that these statements involved amalgamating insights about individuals in these groups. That is, generalization from individuals to groups was accepted, but this idiographic approach virtually prohibited the reverse type of inference.
A second difference was that much of what we did was retrospective. Just as people in the lay public expected psychiatrists and clinical psychologists to “explain” why people ended up with psychological problems by observing what they were like after these problems had developed, we engaged in an approach that can be characterized as “conditioning on the consequence and searching for antecedents” (Dawes, 1993, 2001). In contrast, the work of Meehl began with antecedents (e.g., numerical assessments, test scores, coded variables summarizing past behavior) to see how well they predicted later consequences. As I have argued elsewhere (Dawes, 1993, 2001) conditioning on consequences yields much stronger statistical relationships—hence an illusion of predictability—than does conditioning on antecedents. For starters, when we look at a particular consequence and associate antecedents with it, it is a tautological necessity that those we find are associated with it. In contrast, we can look at antecedents that we expect to be predictive and find that many of the consequences we anticipate are missing. (Could we not also find that many of the antecedents anticipated are missing? Yes, but only if we disciplined ourselves to specify all possible antecedents in advance [ex ante], whereas the more usual search for antecedents—particularly those connected to a consequence for a single individual—is post hoc.) In fact, this approach beginning with consequences can often be little more than “creating a good story”;2 most of us are intelligent and creative enough to do so. Finally, as for the associational base of traditional thinking, see my previous comments about the relationship between symptoms and conditions.

The scientific approach is also different from the traditional one in that we often accept scientific principles and results without fully understanding them. For example, people suffering from psychotic depression who commit suicide often do so when they are getting over the depression, rather than during the depth of depression. Many of us do not know why, although we may hypothesize reasons. In the traditional way of reasoning, we would have to “understand” what is going on with the client to incorporate this regularity, rather than accept it as a simple fact—one we must be aware of if we are to treat psychotic depression.

Let me illustrate this acceptance in another context. Figure 1 summarizes the result of an experiment by McCloskey (1983), who asked undergraduates at Johns Hopkins University to predict the trajectory of a ball that emerged after spiraling through an enclosed area. About half of these undergraduates predicted that the ball would continue spiraling, as illustrated in Figure 1B. Most of the half that predicted the trajectory correctly (as in Figure 1A) had taken Newtonian physics at some point in high school or college. When asked to explain their predictions, they generally referred to “momentum,” but momentum within the ball itself. Momentum does lead to a straight-line trajectory; however, the principle of momentum does not refer to any intrinsic characteristic of

---

2 At a meeting of the Pittsburgh Psychological Association in the late 1980s, I discussed the “good story” problem. I also mentioned that many biases in memory are based on knowing our current state and selectively recalling and interpreting the past to “explain” how we got to this state—most usually in ways consistent with our prior theories of human stability and change. To my surprise, many of the clinicians in the audience agreed with me, stating that they knew that the “story” they helped create for their clients was not historically accurate. The justification was, however, that the clients themselves had created their own narrative story about how they had ended up badly, and that the substitute story the clinicians provided their clients was more mentally healthy than the client’s own stories, and therefore provided them with greater flexibility in the future. I am not sure whether the substitution of one story for another yields or enhances freedom of choice, because the story model beginning with conditioning on the consequences tends to be deterministic. An alternative approach is the “existential” one, that no story explains that clients necessarily do what they do, or feel what they feel. One possible result of this story denial is an enhanced sense of freedom and responsibility for one’s own behavior. To the best of my knowledge, no one has ever conducted a systematic evaluation of the therapeutic effects of story substitution as opposed to story obliteration, let alone a scientific analysis.
objects internal to them, but rather to forces acting outside upon them. Thus, the subjects who had studied Newtonian physics were able to make the correct judgment, even though their understanding of what led to it was faulty. Occasionally, all of us who claim to “accept scientific thinking” draw an inference based on what we believe to be our scientific knowledge without fully understanding its basis. For example, how many people who diet or exercise to avoid obesity could specify the mechanism by which excess weight leads to diabetes? Or the statistical support?

My point in this article is to argue that Meehl should have had something to do with “what we do,” and that in fact, doing our own thing without any constraints from the implication of his work was and is unethical. (At the time, I was in no way prepared to argue that someone else was doing something unethical, but for myself I could reach such a decision. I quit the program and received a degree a few years later in Mathematical Psychology, under the superb direction of the late Professor Clyde Coombs, who was enthusiastic about mentoring work not directly in his area of research but was enthusiastic about all good things in life.)

But how do we get from the “is” of what we were doing and what Meehl had demonstrated to the “ought” of what we should be doing? We all know that we cannot reach an ethical conclusion from purely factual premises (e.g., in this case, that we do this and he does that). Doing so—inferring “ought” from an “is”—was termed by G. E. Moore the “naturalistic fallacy.” This insight of Moore has lead to all sorts of foolishness when interpreted thoughtlessly—from radical (“It’s arbitrary.”) positivism to single act (“What the hell?”) existentialism to solipsistic emotivism (“Ought” means “I like.”) to the current peculiar mixture of deconstructionism and relativism (“It’s all political anyway.”). The point here is that “is” can lead to inferences about “ought” in at least two ways. First, a superordinate “ought” (e.g., that we ought to make best possible predictions for our clients) can combine with an empirical “is” (e.g., that in a wide variety of qualitatively diverse contexts statistical prediction rules outperform clinical combination methods) to yield a specific ought (e.g., that in this context, we should use a statistical prediction rule). Second, what exists constrains what ought to be, or at least what we consider to be reasonable ethical mandates. For example, the fact that people are not capable of flying renders the ethical command that “you ought to fly” a bad one. (There can be disagreements about the “is” here; for example, people might disagree about whether it is possible to “love” someone who is clubbing you so badly that they may create permanent brain injury simply because you are demonstrating in a peaceful manner against some government policy. Ethical thinkers who believe that such love is not possible emphasize the
duty to behave *nevertheless* in a nonviolent and dignified manner. In contrast, Martin Luther King, Jr., apparently believing that such love was possible, emphasized an ethical obligation to try to develop it in the most odious of circumstances.)

I will return at the end of this article to a discussion of the ethical implications of the extensive research indicating that statistical prediction rules (SPRs) outperform clinically based integration of information. First, I discuss three other contexts in which “is” and “ought” combine, and one in which what we know delineates ethical practice—that is, sets up boundaries that we “ought not” to transgress.

First, consider a single “ought” and single “is” from the history of psychiatry and clinical psychology. For years, people were thought to develop schizophrenia as result of being raised by a “schizophrenogenic mother” (or, more rarely, father), who would continually present the “preschizophrenic child” with “double bind” requests or instructions. For example, this pathological parent might repeatedly assert that the child should be completely independent and do what he or she wishes, while suffering public bouts of insomnia, bleeding ulcers, or gout when the child asserts independence by doing something the parent clearly prefers he or she not do. The child had to “opt out” of the close relationship in which the child could *never* really do the right thing. He or she withdrew from unacceptable reality by becoming schizophrenic. What the therapists had to do was to “reparent” the schizophrenic individual, thus becoming what the pathological parent should have been, but was not. That led to some delightful case histories written by at least one such psychotherapist, and some pretty bizarre behavior on the part of a few others.

It was well known that schizophrenia “tended to run in families,” but that could be explained by the hypothesis that family members tend to train each other in how to be schizophrenogenic and how to place their children in double binds. Then in the early 1960s, adoption studies indicated that schizophrenia in children was related to the schizophrenia in the biological parents (and other relatives), but not in the adoptive ones. Of course, it was possible to hypothesize that a tendency to create double binds was passed along in genes, but exactly how that would happen was a bit of a mystery—as opposed to a genetic tendency to develop hallucinations and delusions under appropriate types of stressful situations. (Where the exact nature of the stress is still not known, even from identical twins studies in which one twin later becomes schizophrenic and the other remains apparently normal. See Torrey, Bowler, Taylor, & Gottesman, 1994.) The simple facts following from the adoption studies, combined with the ethical principle that we should do our best to understand the nature of schizophrenia and to treat it, leads us to conclude that the concentration on parenting or re-parenting will not achieve our goals. Combined with the efficacy of such drugs as thior-azine, the adoption studies provided an ethical mandate. In attempting to understand and treat schizophrenia, we should concentrate on biological characteristics. (That does not mean that standard talk therapy cannot help schizophrenics *deal with* their schizophrenic problems, but that we can address neither the etiology nor the amelioration of schizophrenia directly through such therapy.) An ethical clinical psychologist or psychiatrist must accept this inference; to ignore it is unethical.

Or consider the ethical implications of the uniform “dodo bird finding” (“All have won and all must have prizes.”; Luborsky, Singer, & Luborsky, 1975; Smith & Glass, 1977; Stubbs & Bozarth, 1994) about “relationship therapies,” at least as this finding relates to the professional degrees and experience of the therapist. Despite years of trying to find that degrees and experience matter in predicting quality of outcome (often with flawed studies involving few therapists with many clients—even though it is the therapists who were meant to be the objects of study), no relationship has been discovered—although there are occasional claims here and there—at least not a relationship that can serve as a basis for public policy. Now combine this finding with an ethical principle that
states that people supplying a service to others should make clear their qualifications and charge accordingly. If having a degree provides no incremental value as a relationship therapist, then restricting the practice to those with degrees and charging in a manner commensurate with degree attainment is unethical. There are, in contrast to relationship therapies, those that are termed “protocol therapies,” because they follow a well-specified set of procedures, especially those involving cognitive–emotional principles. These therapies, again the cognitive–emotional ones, generally do well, at least when compared to those based on intuition and relationships alone (see Hunsley & Di Giulio, 2002). These, of course, require training in the procedures, which would lead to restrictions in their practice. The principles and procedures of these therapies are, however, not so difficult to understand that their implementation would require advanced degrees (Dawes, 1994).

There are also situations in which multiple (hopefully no conflicting) “shoulds” combine with multiple facts to yield ethical conclusions. One such involves the establishment of sterile needle exchange programs to encourage intravenous drug users to avoid sharing needles (or “the works”) in an attempt to minimize the spread of HIV infection without increasing intravenous drug use. Here, we have two broad ethical mandates. One is to save lives, and the other is to avoid encouraging destructive drug use. Many people originally felt these to be in conflict. Some people, for example, advocated sterile needle exchange programs even if such programs increased drug addiction, whereas others argued that HIV infection was “collateral damage” in the United States “war on drugs.” In contrast, research involving randomized trials both within and across locations indicates quite clearly (Moses, 1994) that sterile needle exchange programs stem the spread of HIV infection, while simultaneously either leaving the rate of intravenous drug use unaffected, or decreasing it (apparently because clients who use the needles are given information about opportunities for treatment programs, when they exist).

Those are the facts. Intuitions that it would be “contradictory” for people who use drugs to care about their health enough to use sterile needles are wrong, as is the intuition that such programs would “send a message” that it is acceptable to use intravenous drugs—despite the fact that users have already indicated through their use an imperviousness to such government-sponsored messages. (The absence of any epidemic of either HIV or of intravenous drug use among diabetics, who have legal access to sterile needles, reinforces these conclusions.) The ethical inference is clear: The government should encourage sterile needle exchange programs. The Clinton administration did, in fact, proclaim that such programs are both “safe and effective” in April 1998, but that did not mean that it was required to support such programs. Instead, Secretary of Health and Welfare, Donna Shalala urged counties and municipalities to implement such programs—as if the HIV retrovirus was too shy to cross local government boundaries.

Empirical results cannot just be combined with superordinate ethical principles to yield subordinate ones, but can also demarcate ethical behavior. The point is that while the “is” does not imply an “ought,” the reverse inference is often valid. For example, if we ought to do something, we can infer that doing it is possible. (The contrapositive form of this inference is that if something is impossible, it cannot yield an ethical mandate.) How does this inference relate to applied psychology?

The relationship is illustrated in Figure 2. Clinicians often point out that research does not tell them exactly what to do, specifically, how to proceed exactly with “this here client.” That is true. What (we believe) we know empirically and theoretically about psychological principles does, however, set bounds in ethical practice. Although it would be ideal to limit principles to the hortatory “do this,” such bounds provide minatory principles about what we ought not to do. (And, in fact, most reprimands, censures, or worse involve violating minatory principles, not failure to live up to hortatory ones.)
The prime example is that of the hypnotic attempt to recover memories in their pristine form, usually memories of alleged childhood sexual abuse—often during Satanic rituals. From everything we know about memory, it is reconstructive. Biases range from overt reliance on cues or questions (“How fast was the car going when it smashed into the bicycle?” versus “What speed was the car going when it hit the bicycle?”), to implicitly biased retrieval of instances from our past, to interpretation of these instances to be consistent with our understanding of our present state and theory-driven beliefs about how we came to it (Pearson, Ross, & Dawes, 1991). Such reconstruction is malleable. Hypnosis can enhance that malleability by the hypnotic subject’s desire to “please” the hypnotist. Thus, careful studies of hypnosis for the retrieval of emotionally neutral stimuli demonstrate that hypnotically enhanced retrieval is more than balanced by false “recall.” (See the National Research Council’s book on “enhancing performance” entitled, Learning, Remembering, Believing; Druckman & Bjork, 1994. For specific studies on the failure of hypnosis in “memory recovery,” see Lynn, Lock, Myers, & Payne, 1997.) Consequently, to attempt to base a therapy on “recovering” memories in their pristine form through hypnosis is simply “out of bounds.” Practitioners should not do it because they cannot do it. Perhaps worst of all is the assertion that there are particular cells or parts of the brain in which these allegedly traumatic memories are stored. We know from positron emission tomography (PET) scans and function magnetic resonance imaging (fMRI) that multiple parts of the brain are involved when people try to recall something, generally the very same areas that are activated when people are exposed to the corresponding visual or verbal information for the first time. For example, the visual cortex is involved in seeing, in recalling visual scenes, and in imagining them. While currently there are attempts to distinguish perception from memory from imagination on a neuro-anatomical basis, they have not yet reached the point where any responsible practitioner could claim with accuracy to be able to distinguish them—especially not through hypnosis!

Now let us consider Paul Meehl’s work on statistical versus actuarial prediction. In his original book (Meehl, 1954/1996), he reviewed articles in which the same input information was presented to clinical experts and used in a statistical analysis to predict important human outcomes (e.g., response to electroshock therapy, success on parole, success in an academic or professional training program). The actuarial prediction always did as well or better than the clinical one. Meehl was extraordinarily careful not to bias the results by failure to “cross-validate” the statistical model, given that commonly used regression techniques automatically maximized prediction for a sample on which they were based. (He also considered unit weights.) Later, Sawyer (1966) extended the analysis to instances in which the clinician might have access to information not in the model; that is to cases in which the information used in the model was a proper subset of the information available to the clinician. Sawyer’s examples were all from psychology, again leading to the conclusion that the statistical models generally outperform clinical judgment.
(Even in those cases where some of the information obtained by the clinician and not used in the model, e.g., from interviews, has predictive validity, it is best incorporated within a statistical model. In other words, if “person and model are to be integrated,” the integration method should be by a model—not a person.) The overall results are summarized by Dawes, Faust, and Meehl (1989), Grove and Meehl (1996), Swets, Dawes, and Moynihan (2000), and Grove et al. (2000). These results—now spanning over 135 studies of predicting important human outcomes—are all consistent with an earlier statement by Meehl that,

There is no controversy in social science which shows such a large body of qualitatively diverse studies coming out so uniformly in the same direction as this one. When you are pushing 90 investigations (now over 135), predicting everything from outcome of football games to the diagnosis of liver disease and when you can hardly come up with a half dozen studies showing even a weak tendency in favor of a clinician, it is time to draw a practical conclusion. (Meehl, 1986, pp. 372–373)

We should draw an ethical conclusion as well. Most of the ambiguous findings are scattered outside the field of psychology, particularly in medicine and business. Even these, however, provide no basis for generalization about when clinical prediction might outperform statistical prediction, and in at least one medical context (predicting 24-hour survival in an intensive care unit) a statistical prediction system that was originally outperformed by clinicians was modified so that it now outperforms the clinicians (see Knaus, Wagner, & Lynn, 1991).

Combining these findings with the superordinate ethical principle that we ought to predict as well as possible, and believing that these findings provide boundaries that we should not cross in practice, we can derive a compelling set of ethical mandates.

First, we ought to use the relevant statistical prediction rule if one is available, rather than rely on clinical judgment in making predictions.

Second, any modification of a rule according to the belief that “My population or time is not exactly equivalent to the population or time in which the SPR was validated,” requires validation. Research has demonstrated that SPRs are quite robust.3

Part of this research yields a third ethical mandate, which is that systematic “improper” rules as well as SPRs constructed to maximize some specific prediction also outperform clinical intuition. The reason is simple. Most of the SPRs, in psychology anyway, consist of linear models, and the weights used in such models—provided they are in the right direction (which is usually easy to assess)—are not very important. That is, small variations, or even large variations, in weights yield similar outputs—provided that weights are applied systematically. (Again, the proviso is that predictor variables not be weighted in the wrong direction.) That is what Dawes and Corrigan (1974) showed by constructing “random linear models,” which outperform clinicians. Unit weights also often perform remarkably well, a result first noted by Wilks (1938).

Thus, there is no ethical justification for using intuition rather than a SPR to make a prediction. That is often not a popular or compelling conclusion, any more than the one

3 People who object bring up specific instances such as “Well surely, if the applicant’s mother had just died you wouldn’t want to treat her test score like that of any other applicant. Wouldn’t you want to add a few points?” One answer is to note that applicants are often in a competitive situation, and adding a few points to the score of the applicant whose mother died is logically identical to subtracting those from every other applicant, who is thereby “punished” for having a mother who didn’t just die, and changing the cut-score accordingly. Moreover, if such modifications—often termed “broken leg exceptions” following Meehl (1954)—are individually predictive, then they would be predictive in the aggregate, but the research shows that modifying SPRs through the clinical injection of such “special circumstances” yields inferior prediction, not better prediction.
about the credentials and experience of psychotherapists who engage in relationship therapy. It just happens to be true.

Moreover, it is possible to construct one’s own SPR for a particular situation by collecting a number of observations before making predictive judgments. (And if there are not enough observations to make predictive judgments, perhaps they should not be made at all.) A particularly compelling finding of Dana and Dawes (2004) is that many composites formed by simple correlation coefficients are very good when applied to data sets from which samples are drawn, where unlike regression weights, correlation values are not dependent on the other variables considered in the prediction and the relationships (covariances) among the variables. Wainer (1978) and Grove (2002) specify the (broad) class of situations in which unit weights do well. Correlation weights often do better, without having the problem of instability due to the intercorrelations among predictors.

Finally, SPRs have a particular virtue that human intuition does not; by specifying how well a prediction can be made, they automatically specify how badly it is made. Knowing the limits of predictability—and sharing them with clients, whether they are individuals or organizations—can also be justified as an ethical mandate, along the lines presented in this article. Ironically, the automatic specification of how poorly a prediction is made yields a temptation to believe that because it could, hypothetically anyway, be made better, the relevant SPR should be abandoned in preference to some unproven method—most often our own intuition, which we have found to be worse. As on other occasions, ethical behavior involves resisting temptation.

References


