*Journal of Counseling Psychology*
Vol. 4, No. 4, 1957

# When Shall We Use Our Heads Instead of the Formula?

Paul E. Meehl[1]
*University of Minnesota*

My title question, "When should we use our heads instead of the formula?" is not rhetorical. I am sincerely asking what I see as an important question. I find the two extreme answers to this question, namely, "Always" and "Never," equally unacceptable. But to formulate a satisfactory answer upon the present evidence seems extraordinarily difficult.

I put the question in the practical clinical context. This is where Sarbin put it in his pioneering study 14 years ago, and this is where it belongs. Some critics of my book (5) have repudiated the whole question by saying that, always and necessarily, we use *both* our heads and the formula. No, we do not. In research, we use both; the best clinical research involves a shuttling back and forth between clever, creative speculation and subsequent statistical testing of empirical derivations therefrom. So far as I am aware, nobody has ever denied this. Even the arch-actuary George Lundberg approved of the clinician as hypothesis-maker. In research one cannot design experiments or concoct theories without using his head, and he cannot test them rigorously without using a formula. This is so obvious that I am surprised to find that people will waste time in discussing it. The clinical-statistical issue can hardly be stated so as to make sense in the research context, and I should have thought it clear that a meaningful issue can be raised only in the context of daily clinical activity.

In the clinical context, on the other hand, the question is sensible and of great practical importance. Here we have the working clinician or administrator, faced with the necessity to make a decision at *this* moment in time, regarding *this* particular patient. He knows that his evidence is inadequate. He can think of several research projects which, *had* they been done already, would be helpful to him in deciding the present case. If he is research-oriented he may even make a note of these research ideas and later carry them out or persuade someone else to do so. But none of that helps him *now*. He is in a sort of Kierkegaardian existential predicament, because he has to act. As Joe Zubin kept repeating when I last tangled with him on this subject, "Every clinical decision is a *Willensakt*." And so it is; but the question remains, how do we make our *Willensakts* as rational as possible upon limited information? *What clinician X knows today* and *what he could find out by research in ten years* are two very different things.

The question, "When shall we use our heads instead of the formula?" presupposes that we are about to make a clinical decision at a given point in time, and must base it upon what is known to us at that moment. In that context, the question makes perfectly good sense. It is silly to answer it by saying amicably, "We use both methods, they go hand in hand." If the formula and your head invariably yield the same predictions about individuals, you should quit using the more costly one because it is not adding anything. If they don't always yield the same prediction—and they clearly don't, as a matter of empirical fact—then you obviously can't "use both," because you cannot predict in op-

[1]Presented at the 1956 Convention of the American Psychological Association, Chicago. This will also appear in a forthcoming book: Herbert Feigl, Michael Scriven, and Grover Maxwell (Eds.), *Concepts, Theories, and the Mind-Body Problem*, Vol. II of Minnesota Studies in the Philosophy of Science. Minneapolis, Minnesota: University of Minnesota Press, 1958.

posite ways for the same case. If one says then, "Well, by 'using both,' I mean that we follow the formula except on special occasions," the problem becomes how to identify the proper sub-set of occasions. And this of course amounts to the very question I am putting. For example, does the formula tell us "Here, use your heads," or do we rely on our heads to tell us this, thus countermanding the formula?

## The Pragmatic Decision Problem Stated

Most decisions in current practice do not pose this problem because no formula exists. Sometimes there is no formula because the prediction problem is too open-ended, as in dream analysis; sometimes the very categorizing of the raw observations involves Gestalted stimulus equivalences for which the laws are unknown, and hence cannot be mathematically formulated (although the clinician himself exemplifies these laws and can therefore "utilize" them); in still other cases there is no formula because nobody has bothered to make one. In any of these three circumstances, we use our heads because there isn't anything else to use. This presumably will be true of many special prediction situations for years to come. The logical analysis of the first two situations—open-endedness and unknown psychological laws—is a fascinating subject in its own right, especially in relation to psychotherapy. But since our original question implies that a formula does exist, we will say no more about that subject here.

Suppose then that we have a prediction equation (or an actuarial table) which has been satisfactorily cross-validated. Let us say that it predicts with some accuracy which patients will respond well to intensive outpatient therapy in our VA clinic. We are forced to make such predictions because our staff-patient ratio physically precludes offering intensive treatment to all cases; also we know that a minority, such as certain latent schizophrenias, react adversely and even dangerously. The equation uses both psychometric and non-psychometric data. It may include what the Cornell workers called "Stop" items—items given such a huge weight that when present they override any combination of the remaining factors. It may be highly patterned, taking account of verified interaction effects.

So here is veteran Jones, whose case is under consideration at therapy staff. The equation takes such facts as his Rorschach $F+$, his Multiphasic code, his divorce, his age, his 40 per cent service-connection, and grinds out a probability of .75 of "good response to therapy." (The logicians and theoretical statisticians are still arguing over the precise meaning of this number as applied to Jones. But we are safe in saying, "If you accept patients from this population who have this score, you will be right 3 times in 4.") Here is Jones. We want to do what is best for him. We don't *know for sure,* and we can't, by any method, actuarial or otherwise. We act on the probabilities, as everyone does who chooses a career, takes a wife, bets on a horse, or brings a lawsuit. (If you object, as some of the more cloud-headed clinikers do, to acting on "mere probabilities," you will have to shut up shop, because probabilities are all you'll ever get.)

But now the social worker tells us that Jones, age 40, said at intake that his mother sent him in. The psychology trainee describes blocking and a bad $F-$ on Rorschach VII; the psychiatrist adds his comments, and pretty soon we are concluding that Jones has a very severe problem with mother-figures. Since our only available therapist is Frau Dr. Schleswig-Holstein, who would traumatize anybody even without a mother-problem, we begin to vacillate. The formula gives us odds of 3 to 1 on Jones; these further facts, not in the equation, raise doubts in our minds. What shall we do?

## Importance of 'Special Cases'

In my little book on this subject, I gave an example which makes it too easy (**5,** p. 24). If a sociologist were predicting whether Professor X would go to the movies on a certain night, he might have an equa-

tion involving age, academic specialty, and introversion score. The equation might yield a probability of .90 that Professor X goes to the movie tonight. But if the family doctor announced that Professor X had just broken his leg, no sensible sociologist would stick with the equation. Why didn't the factor of "broken leg" appear in the formula? Because broken legs are very rare, and in the sociologist's entire sample of 500 criterion cases plus 250 cross-validating cases, he did not come upon a single instance of it. He uses the broken leg datum confidently, because "broken leg" is a subclass of a larger class we may crudely denote as "relatively immobilizing illness or injury," and movie-attending is a subclass of a larger class of "actions requiring moderate mobility." There is a universally recognized "subjective experience table" which cuts across sociological and theatrical categories, and the probabilities are so close to zero that not even a sociologist feels an urge to tabulate them! (That this is the correct analysis of matters can be easily seen if we ask what our sociologist would do if he were in a strange culture and had seen *even a few* legs in casts at the movies?)

I suppose only the most anal of actuaries would be reluctant to abandon the equation in the broken leg case, on the ground that we were unable to cite actual statistical support for the generalization: "People with broken legs don't attend movies." But clinicians should beware of overdoing the broken leg analogy. There are at least four aspects of the broken leg case which are very different from the usual "psychodynamic" reversal of an actuarial prediction. First, a broken leg is a pretty objective fact, determinable with high accuracy, if you care to take the trouble; secondly, its correlation with relative immobilization is near-perfect, based on a huge *N*, and attested by all sane men regardless of race, creed, color, or what school granted them the doctorate; thirdly, interaction effects are conspicuously lacking—the immobilization phenomenon cuts neatly across the other categories under study; fourthly, the prediction is mediated without use of any doubtful theory, being either purely taxonomic or based upon such low-level theory as can be provided by skeletal mechanics and common sense. The same cannot be said of such an inference as "Patient Jones has an unconscious problem with mother-figures, and male patients with such problems will not react well in intensive therapy with Frau Dr. Schleswig-Holstein."

### Theoretical Derivation of Novel Patterns

When the physicists exploded the first atomic bomb, they had predicted a novel occurrence by theoretical methods. No actuarial table, based upon thousands of combinations of chemicals, would have led to this prediction. But these kinds of theoretical derivations in the developed sciences involve combining rigorously formulated theories with exact knowledge of the state of the particular system, neither of which we have in clinical psychology. Yet we must do justice to the basic *logical* claim of our clinician. I want to stress that he is not in the untenable position of denying the actuarial data. He freely admits that 75 per cent of patients having Jones' formula score are good bets for therapy. But he says that Jones belongs to the other 25 per cent, and therefore thinks we can avoid one of our formula's mis-predictions by countermanding the formula in this case. There is nothing intrinsically wrong with this suggestion. Perhaps the clinician *can* identify a subclass of patients within the class having Jones' actuarial attributes, for which the success rate is less than .5. This would be perfectly compatible with the over-all actuarial data, provided the clinician doesn't claim it too often.

At this point the actuary, a straightforward fellow, proposes that we tabulate the new signs mentioned in staff conference as indicating this subclass before proceeding further. Here we again reduce our clinician to a hypothesis-suggestor, and seem to put the current prediction problem back on an actuarial basis. But wait. Are we really prepared to detail someone to do such "case-oriented" research every

time a clinical prediction is made? Actually it is impossible. It would require a superfile of punch-cards of colossal $N$ to be available in each clinic, and several major staff doing nothing but running case-oriented minor studies while clinical conferences went into recess pending the outcomes.

However, this is a "practical" objection. Suppose we circumvent it somehow, so that when a sign or pattern is used clinically to support a counter-actuarial prediction, we can proceed immediately to subject the sign to actuarial test on our clinic files. There are serious difficulties even so. Unless the several staff who produced these records had in mind all of the signs that anybody subsequently brings up, we have no assurance that they were looked for or noted. Anyone who has done file research knows the frustration of having no basis for deciding when the lack of mention of a symptom indicates its absence. But even ignoring this factor, what if we find only 3 cases in the files who show the pattern? *Any split* among these 3 cases as to therapy outcome is statistically compatible with a wide range of parameter values. We can neither confirm nor refute, at any respectable confidence level, our clinician's claim that this pattern brings the success-probability from .75 to some value under .5 (he doesn't say how far under).

Here the statistician throws up his hands in despair. What, he asks, can you do with a clinician who wants to countermand a known probability of .75 by claiming a subclass probability which we cannot estimate reliably? And, of course, one wonders how many thousands of patients the clinician has seen, to have accumulated a larger sample of the rare configuration. He also is subject to sampling errors, isn't he?

## Non-frequentist Probability and Rational Action

This brings us to the crux of the matter. Does the clinician need to have seen *any* cases of "mother-sent-me-in" and Card VII blockage who were treated by female therapists? Here we run into a philosophical issue about the nature of probability. Many logicians (including notably Carnap, Kneale, Sellars, and most of the British school) reject the view (widely held among applied statisticians) that *probability is always frequency.* Carnap speaks of "inductive probability," by which he means the logical support given to a hypothesis by evidence. We use this kind of probability constantly, both in science and in daily life. No one knows how to compute it exactly, except for very simple worlds described by artificial languages. Even so, we cannot get along without it. So our clinician believes that he has inductive evidence from many different sources, on different populations, partly actuarial, partly experimental, partly anecdotal, that there is such a psychological structure as a "mother-surrogate problem." He adduces indirect evidence for the construct validity (1) of Rorschach Card VII reactions. I am not here considering the actual scientific merits of such claims in the clinical field, on which dispute still continues. But I think it important for us to understand the methodological character of the clinician's rebuttal. If Carnap and some of his fellow-logicians are right, the idea that *relative frequency* and *probability* are synonymous is a philosophical mistake.

Of course there is an implicit future reference to frequency even in this kind of inductive argument. Carnap identifies inductive probability with the betting odds which a reasonable man should accept. I take this to mean that if the clinician decided repeatedly on the basis of what he thought were high inductive probabilities, and we found him to be wrong most of the time, then he was presumably making erroneous estimates of his inductive probabilities. The claim of a high inductive probability implies an expectation of being right; in the long run, he who (correctly) bets odds of 7:3 will be able to point to a hit-rate of 70 per cent. But this *future* reference to success-frequency is not the same as the *present evidence for* a hypothesis. This seems a difficult point for people to see. As a member of a jury, you might be willing to bet 9 to 1 odds on the prisoner's

guilt, and this might be rational of you; yet no calculation of frequencies constituted your inductive support in the present instance. The class of hypotheses where you have assigned an inductive probability of .9 should "pan out" 90 per cent of the time. But the assignment of that inductive probability to each hypothesis need not itself have been done by frequency methods. If we run a long series on Sherlock Holmes, and find that 95 per cent of his "reconstructions" of crimes turn out to be valid, our confidence in his guesses is good *in part just because they are his.* Yet do we wish to maintain that a rational man, ignorant of these statistics, could form no "probable opinion" about a *particular* Holmesian hypothesis based on the evidence available? I cannot think anyone wants to maintain this.

The philosophical recognition of a nonfrequency inductive probability does not help much to solve our practical problem. No one has quantified this kind of probability (which is one reason why Fisher rejected it as useless for scientific purposes). Many logicians doubt that it can be quantified, even in principle. What then are we to say? The clinician thinks he has "high" (How high? Who knows?) inductive support for his particular theory about Jones. He thinks it is so high that we are rationally justified in assigning Jones to the 25 per cent class permitted by the formula. The actuary doubts this, and the data do not allow a sufficiently sensitive statistical test. Whom do we follow?

## Monitoring the Clinician

Well, the actuary is not quite done yet. He has been surreptitiously spying upon the clinician for, lo, these many years. The mean old scoundrel has kept a record of the clinician's predictions. What does he find, when he treats the clinician as an empty decision-maker, ignoring the inductive logic going on inside him? Let me bring you up to date on the empirical evidence. As of today, there are 27 empirical studies in the literature which make some meaningful comparison between the predictive success of the clinician and the statistician. The predictive domains include: success in academic or military training, recidivism and parole violation, recovery from psychosis, (concurrent) personality description, and outcome of psychotherapy. Of these 27 studies, 17 show a definite superiority for the statistical method; 10 show the methods to be of about equal efficiency; none of them show the clinician predicting better. I have reservations about some of these studies; I do not believe they are optimally designed to exhibit the clinician at his best; but I submit that it is high time that those who are so sure that the "right kind of study" will exhibit the clinician's prowess, should *do* this right kind of study and back up their claim with evidence. Furthermore, *a good deal of routine clinical prediction is going on all over the country in which the data available, and the intensity of clinical contact, are not materially different from that in the published comparisons.* It is highly probable that current predictive methods are costly to taxpayers and harmful to the welfare of patients.

Lacking quantification of inductive probability, we have no choice but to examine the clinician's success-rate. One would hope that the rule-of-thumb assessment of inductive probability is not utterly unreliable. The indicated research step is therefore obvious: We persuade the clinician to state the odds, or somehow rate his "confidence," in his day-by-day decisions. Even if he tends over-all to be wrong when countermanding the actuary, he may still tend to be systematically right for a high-confidence sub-set of his predictions. Once having proved this, we could thereafter countermand the formula in cases where the clinician expresses high confidence in his head. It is likely that studies in a great diversity of domains will be required before useful generalizations can be made.

In the meantime, we are all continuing to make predictions. I think it is safe to say, on the present evidence, that we are not as good as we thought we were. The development of powerful actuarial methods could

today proceed more rapidly than ever before. Both theoretical and empirical considerations suggest that we would be well advised to concentrate effort on improving our actuarial techniques rather than on the calibration of each clinician for each of a large number of different prediction problems. How should we meanwhile be making our decisions? Shall we use our heads, or shall we follow the formula? Mostly we will use our heads, because there just isn't any formula, but suppose we have a formula, and a case comes along in which it disagrees with our heads? Shall we then use our heads? I would say, yes—provided the psychological situation is as clear as a broken leg; otherwise, very, *very* seldom.

*Received August 1, 1957.*

## References

1. Cronbach, L. J. & Meehl, P. E. Construct validity in psychological tests. *Psychol. Bull.*, 1955, **52**, 281-302.
2. Humphreys, L. G., McArthur, C. C., Meehl, P. E., Sanford, N., & Zubin, J. Clinical versus actuarial prediction. *Proceedings of the 1955 Invitational Conference on Testing Problems*, pp. 91-141.
3. McArthur, C. C. Analyzing the clinical process. *J. counsel. Psychol.*, 1954, **1**, 203-207.
4. McArthur, C. C., Meehl, P. E., & Tiedeman, D. V. Symposium on clinical and statistical prediction. *J. counsel. Psychol.*, 1956, **3**, 163-173.
5. Meehl, P. E. *Clinical versus statistical prediction: a theoretical analysis and a review of the evidence* Minneapolis: Univer. of Minn. Press, 1954
6. Meehl, P. E. Comment on Analyzing the Clinical process. *J. counsel Psychol.*, 1954, **1**, 207-208.
7. Meehl, P. E. Wanted—a good cookbook. *Amer. Psychologist*, 1956, **11**, 263-272.
8. Meehl, P. E. & Rosen, A. Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychol. Bull.*, 1955, **52**, 194-216.