Acta Psychologica 56 (1984) 93–109
North-Holland

METHODOLOGICAL DECISION RULES AS RESEARCH POLICIES: A BETTING RECONSTRUCTION OF EMPIRICAL RESEARCH

Willem K.B. HOFSTEE *

University of Groningen, The Netherlands

A betting model of empirical research is described. The model requires that opposing parties reach agreement on an operationalization and specify their predictions in terms of a probability distribution over possible research outcomes. Proper decision rules are used to decide on the amounts of reputation that are gained and lost upon observing the data. It is argued that adoption of the model would lead to less trivial research, less selective publication, and a more liberal attitude towards experimental design. The betting model is contrasted with several other methodologies. In these comparisons, methodological models are viewed as policies, i.e., sets of rules which lead to certain predictable consequences if rational individuals exploit these rules to their own advantage. All comparisons reveal that the other models have relatively undesirable properties from this point of view. In discussing criticisms of the betting model, it appears that the model is also not completely water-proof from a policy point of view when participants do not wish to maximize their subjectively expected reputation. Other criticisms are discussed and are found wanting.

Bets enter quite naturally into scientific discussions. Some years ago, an American colleague and I were discussing handwriting analysis. We were in general agreement that the usefulness of handwritings for making predictions about persons is very limited at best. As a devil's advocate, however, I asserted that I could "predict" a person's nationality from his or her handwriting. My colleague, who probably suspected me of entertaining ideas about national character, disagreed strongly with my assertion, and we decided to bet upon it. Those who have corresponded in handwriting with foreigners will not be surprised that I won this bet.

* The author is indebted to Lewis R. Goldberg and Sara Lichtenstein for their sympathetic but thorough criticisms of earlier versions.

Author's address: W.K.B. Hofstee, Vakgroep Persoonlijkheidsleer, Subfaculteit Psychologie, Grote Markt 31/32, 9712 HV Groningen, The Netherlands.

In the following, a betting model of empirical discussions will be presented, the core of which consists of decision rules for scientific bets. Subsequently, I shall argue that the betting model, viewed as a methodological policy, leads to desirable consequences and should be preferred over certain other methodological approaches. Finally, some reservations about, and criticisms of, the betting model will be discussed.

The betting model

Assume that two or more parties (e.g., persons, schools of thought) are in disagreement over an issue, and agree in principle to resolve their dispute empirically instead of, for example, through violence or semantic analysis. This is to say that their positions on the issue are held to be commensurable in the sense that the parties are willing in principle to accept some common empirical ground.

Operationalizations

The parties will then look for a mutually acceptable operationalization of the concepts that are central in the discussion. In the above example, "handwriting" was operationalized by having subjects copy a Latin sentence in their own handwriting. In other cases, it may not be so easy to reach agreement upon an operationalization, and it is not assumed here that such attempts will always be successful. If no common operationalization can be found, however, then empirical research serves no function in the discussion. Therefore, such cases are outside the scope of the present methodological model.

In still other cases, the operationalization phase is sufficient to resolve the dispute. For example, in discussing whether a particular alternative medical treatment is beneficial, the parties may at first have different operationalizations in mind. "Beneficial" may mean to one party that patients are satisfied with the treatment, whereas to the other it may mean that their life expectancy is increased. As soon as these operationalizations are made explicit, the parties may agree that the treatment is probably beneficial under the one, but not under the other, operationalization. The scope of the betting model is restricted to those cases in which substantive disagreement remains after agreement upon an operationalization is reached.

Predictions

The next phase in the betting scenario is that the parties specify their predictions. In the general case, a well-formed prediction should have the shape of a probability distribution, reflecting, for example, a party's probability that the percentage of nationalities correctly identified on the basis of handwriting will be 51%–60%, 61%–70%, and so on.

Since the parties have differing predictions, and since there is supposedly no objective way of determining whether these predictions are a priori correct (if such a criterion were available, the discussion would be senseless), the distributions should be conceived as personal or subjective probability distributions. The personal distribution $p(x)$ with respect to the observed outcome x may be derived from a prior (Bayes') distribution $b(\theta)$ with respect to the population parameter θ, as follows:

$$p(x) = \int p(x|\theta) b(\theta) d\theta, \tag{1}$$

in which $p(x|\theta)$ is the model distribution. Assuming that the parties can easily agree upon a model distribution, their difference of opinion can be traced back to a difference between prior distributions $b(\theta)$. In this respect the betting model is Bayesian. It is non-Bayesian or rather a-Bayesian, however, in the sense that posterior beliefs have no place in the model.

Decision rules

In the handwriting bet, we took statistical significance as a criterion. Objections against statistical hypothesis testing (e.g., Morrison and Henkel 1970), however, become magnified in the betting situation. A party could enhance its probability of winning simply by decreasing the sample size (in case the null hypothesis is entertained) or by increasing it (in case the alternative hypothesis is entertained). In practice, this property of the significance rule would therefore lead to a bargaining process in which nonscientific elements could easily enter. Secondly, the significance rule would require that one party entertain the null hypothesis. This is a very restrictive requirement, since in most cases the null hypothesis represents an unrealistic position. Thirdly, and more generally, classical statistics can admit only distributions $p(x)$ that can be derived from a point hypothesis, not a more or less vague prior distribution. The implication is that parties are subjectively certain about the location of the population parameter θ. That implication is unrealistic.

On the basis of these considerations, the following decision rule might spring to mind, and has in fact been proposed independently by Hofstee (1977) and Pitz (1978):

$$v_{AB} = \frac{p_A(X) - p_B(X)}{p_A(X) + p_B(X)}, \tag{2}$$

where v_{AB} is what A receives from B, in some arbitrary unit; $p_A(X)$ is the predictive probability assigned by A to the outcome X which was in fact observed; and $p_B(X)$ is B's predictive probability of the actual outcome.

There remains a problem, however, with respect to this rule and several other rules that could be constructed. As soon as A has specified a distribution $p_A(x)$, it would be advantageous to B to submit a distribution $r_B(x) \neq p_B(x)$, taking A's predictions into account. This implication is undesirable, since A would in turn wish to respecify his or her predictions, taking $r_B(x)$ into account, et cetera. The decision rule is thus "improper" in the sense that it encourages distortion (cf. De Finetti 1965; Van Naerssen 1961; Shuford et al. 1966).

The defining characteristic for the (strict) properness of a decision rule can be formulated as follows:

$$SEV = \max \text{ iff } r(x) = p(x),$$

that is, the rule should be such that a party maximizes its subjectively expected value SEV only by submitting an $r(x)$ which is identical to its $p(x)$.

To ensure that parties are motivated to bet, the rule should be such that their subjectively expected value is positive, and strictly positive if and only if their predictions differ. Thus a second requirement is that $SEV \geqq 0$, provided that $r(x) = p(x)$, and $SEV > 0$ if moreover $p_A(x) \neq p_B(x)$.

A third requirement that might be posed is symmetry of the decision rule, i.e., $SEV_A = SEV_B$, provided that $r(x) = p(x)$ for both parties. If a rule is symmetric for all sample sizes, a party cannot enhance its subjectively expected pay-off at the cost of the other party's *SEV*. Since the interests of the parties are thus parallel, they may decide upon an optimal sample size in harmony, taking research costs into account.

The following decision rule may be shown to satisfy all these requirements:

$$v_{AB} = p_A(X) - p_B(X) - \tfrac{1}{2}\sum_j p_A^2(x_j) - p_B^2(x_j), \tag{3}$$

in which the sum is taken over all possible outcomes x_j. If the set of possible outcomes is dichotomized such that $x = \{X, \bar{X}\}$, $\bar{X}$ being the outcome that is *not* observed, the rule simplifies to:

$$v_{AB} = p_B^2(\bar{X}) - p_A^2(\bar{X}). \tag{4}$$

It has been shown by Nevels (1980) that eq. (4) is the only decision rule that satisfies all the above requirements in the binary case.

As an illustration, take the case in which *A* believes with probability 0.9 that he will get at least 75% of the nationalities right (x_1) in the handwriting bet, and *B* believes with probability 0.6 that *A* will not (x_2):

	x_1	x_2
p_A	0.90	0.10
p_B	0.40	0.60

The payoff-matrix, according to eq. (4), is as follows:

	x_1	x_2
v_{AB}	0.35	−0.65
v_{BA}	−0.35	0.65

It may be verified that $SEV_A = SEV_B = 0.25$. Note that the decision rule can also accommodate cases in which *A* and *B* agree, but one is more certain than the other (e.g., 0.90 versus 0.70).

A final problem concerns the nature of the gains and losses. The use of money or other real-life values would introduce difficulties which will be discussed in the section on criticisms of the betting model. An appropriate designation of what is gained or lost in empirical-scientific discourse might be reputation: a person who consistently predicts better than others sees his or her reputation enhanced at the cost of others' reputations. Naturally, an opponent should have at least some reputation – whether scientific, as in basic science, or social, as in applied science – to be eligible for a bet. If more than two parties are involved, the number of bets and the subjectively expected gains rise accordingly. Some central authority would have to keep score. For the area of empirical research, this reputation score may be found less objectionable than counts of publications or citations.

Systematic institution of the model would give rise to an economy with desirable and undesirable side effects. However, the present state of the scienficic enterprise may also be described as an economy, and there is ample evidence of undesirable side effects of the rules – implicit and explicit – by which it is now governed. The next section discusses some of the changes that would take place if the betting model would be adopted.

Implications of the model

Methodological models are usually evaluated in terms of their descriptive accuracy or their normative acceptability. Descriptive accuracy is hardly an appropriate criterion, however, since it cannot be assumed that the current scientific state of affairs is desirable. Normative acceptability may be conceived in two rather different ways: a methodology may be judged either as a program or as a policy. In the former conception, the criterion is idealistic; a methodology is judged on whether it professes the right values, such as objectivity, impartiality, logical coherence, and the like. The idealistic conception of methodological programs, however, has no means of ascertaining whether these values are implemented. Furthermore, it cannot prevent a program from serving as an ideology masking an undesirable state of affairs.

A pragmatic conception of evaluation will be adopted here, in which a methodological model is viewed as a policy to be judged on the basis of its expected effects, whether intended or unintended, rather than its descriptive accuracy or its intrinsic nobility. Admittedly, the betting model has little to offer in these latter respects. Certain characteristics of current scientific practice, to be discussed below, cannot be reconstructed from the model, and instead of postulating elevated values it appeals to an "economic" motive, i.e., enhancing one's reputation. I shall argue, however, that such properties of a model are not necessarily shortcomings.

Implications of the model with respect to relevance versus triviality of research, selective publication, control of artifacts, and experimental design will be sketched briefly.

The triviality problem

A major pretension of the betting model is that it prevents trivial research. Inevitably this presumption is partly a matter of definition, but the definition in its turn is not trivial.

The subjectively expected amount of reputation to be gained in a bet, under the decision rule presented above, is a function of the predictive probability distributions $p_A(x)$ and $p_B(x)$ of parties A and B. For both parties, this subjective expectation is equal to:

$$\tfrac{1}{2}\sum_{j}\left\{p_A(x_j)-p_B(x_j)\right\}^2. \tag{5}$$

The expectation is zero when $p_A(x)=p_B(x)$ and maximal when $p_A(x)$ and $p_B(x)$ have no overlap; generally, a bet is more attractive as the predictions are further apart. The betting model thus favors research on issues that are controversial in the above sense.

It is a widely held view that the social and behavioral sciences tend to indulge in proving what everyone already knows. To the extent that this view is a fallacy of hindsight, the betting approach would eradicate the illusion by requiring predictions before the fact instead of after. To the extent that the perception is veridical, however, the model provides an equally radical solution since it eliminates the motive for trivial research.

The betting model is an obstacle to trivial research in still another respect. The use of trivial, externally invalid operationalizations of important concepts is a familiar problem to the social and behavioral sciences. The problem is caused at least in part by idiosyncrasies that investigators are at liberty to exhibit. Moreover, there is no obvious solution to the problem since it is hardly possible to devise rules for finding creative and valid operationalizations. The betting approach provides a safeguard against triviality by requiring that at least two parties who are in substantive disagreement and have a stake in the issue accept the operationalization.

More generally, the betting approach reflects a conception of methodology or even policy which deviates from most other conceptions. The debate with respect to the regulation of behavior (of scientists or of "citizens" in general) is often between two principles, one being the appeal to the individual conscience as the prime means to bring about desirable behavior, the other the institution of an authority – a scientific forum or a state – to enforce regulation. The emphasis in the present conception, however, is neither upon the individual conscience nor upon the superseding authority, but upon social bargaining processes. Policy is viewed as a minimal set of rules under which bargaining on the basis of self-interest leads to desirable external effects, in this case, to nontrivial research. Of course the emphasis is relative: proper bargaining presupposes both a certain minimal appeal to the individual conscience and a minimal authority.

Selective publication

There is a widespread suspicion that statistically significant results are overrepresented in the social and behavioral journals. The reasoning (Cohen 1962) behind this suspicion is as follows: Upon noting the sizes of the observed experimental effects or correlations and the sample sizes that were used in the investigations, it can be argued that fewer statistically significant effects would be expected than are reported in the literature. The inference is that selective effects favoring statistically significant results are operative, and that published results are a biased sample of observed results.

The inference is strongly supported by consideration of the rewards and sanctions that are implied in the classical statistical methodology. There is a clear asymmetry, according to that methodological model, between significance and insignificance. Finding a statistically significant outcome enables one to reject the null hypothesis, according to the convention embodied in the statistical methodology; observing insignificance, however, means that the results are inconclusive. Authors and editors are therefore under pressure to report and publish significant findings.

More fundamentally, this undesirable state of affairs is caused by the use of classical statistics for inductive purposes. Logically speaking, classical statistics is a method for making predictions conditional upon a hypothesis, not for making inductions such as rejection of the null hypothesis. These inductions can only rest upon a supplementary convention which is outside the logical framework of this methodology. It may be possible to design conventions that would have less undesirable external effects. The radical solution, however, would be to refrain from inductive application of classical statistics.

No conventions permitting inductive interpretation are incorporated in the betting model. On the basis of this and other properties I shall argue that the model does not elicit selective publication.

Under the model, the aim of initial publications about an empirical issue is to challenge potential opponents. The author expects a gain in reputation if the betting offer is accepted and the prediction put to the test. The author may have carried out empirical research prior to publishing an assertion. It would be quite irrational, however, to capitalize on chance in formulating such a prediction. In the first place, empirical results would hardly be accepted as being in any way conclusive, since potential opponents have no prior commitment to the operationalizations, the design, and the analysis; therefore, there is no premium upon presenting impressive results, whether biased or not. In fact, the presentation of empirical findings at this stage would serve only secondary purposes like illustrating the kind of research that the author has in mind. Secondly, if the author had any suspicion whatsoever that the findings were biased, it would be disadvantageous to report such findings. For under the proper scoring rule presented above, the publication of a prediction that does not reflect one's convictions entails a decrement in subjectively expected value. Not only does the model not elicit selective publication, it even urges the investigator to be critical of results that are too good to be true.

Subsequent publications on an issue would ideally be copublications by two or more opponents, reporting the outcomes of crucial experiments. The internal purpose of these publications would be to record the reputation gains and losses of the participants. The external effect would be that others can make up their minds with respect to the substantive issue. The reason why selective publication at this stage is ruled out is that a prior commitment to publicize the results is an integral part of the bargain.

It may well be objected that a methodological state of affairs in which the scientific community would embrace the betting model is difficult to envisage, in view of the stern consequences with respect to publication habits. Certainly there are few if any examples of adversarial copublications, fascinating though the prospect may be. However, a model may be influential without being universally adopted: It makes a difference whether researchers and their public conceive empirical science as a sequence of incommensurable paradigms that pop up and fade away in the manner of fashions, as an individual hunt for statistical significance, or as a rationally competitive enterprise. The betting model emphasizes certain tendencies already implied in scientific practice – such as anticipating others' objections against an operationalization, design, or analysis, avoiding predictions that do not reflect one's beliefs, realizing that exploratory research tends to be unconvincing, and preferring to test competing hypotheses over single hypothesis testing – at the cost of other, presumably less productive, tendencies.

Control of artifacts and experimental design

In an important paper, Campbell (1969) has argued that artifacts should be controlled only if empirically plausible, not if merely possible. If a potential artifact, for example, pretest sensitization, has been repeatedly observed to be absent, it would be a waste of money or statistical power (depending on which of the two is held constant) to keep controlling that artifact.

In the context of the betting model, assertions about the plausibility of an artifact would be treated in the same way as assertions about substantive effects, namely as personal predictive probabilities. Elaboration of this interpretation leads to a revision of Campbell's position, since two parties are involved.

If both parties who disagree over a substantive issue, find the artifact implausible they will decide not to control since to do so would be uneconomical. If there is disagreement about its plausibility, the artifact issue will be put to the test, which means installing controls. Interestingly however, if both parties find the artifact plausible, there is no incentive to investigate or control it; it would simply be taken into account. For example, if parties agree about the expected direction and size of a retest effect, that effect will be subtracted from the experimental effect. Thus the revision of Campbell's proposition is that artifacts should be controlled only if they are controversial, rather than if they are plausible.

In conclusion, the betting approach retains the liberal and pragmatic attitude towards the control of artifacts that was advocated by Campbell. The consequences of that attitude, however, are different.

The issue of experimental control may be phrased in still another way. Under the betting model, the control condition – if any – should be conceived as the implementation of some opponent's prediction. For example, if someone asserts that a form of psychotherapy has a success rate of 80%, an opponent, instead of denying the proponent's claim, may predict that non-treatment will have the same success rate. If the proponents disagree with the latter prediction, as they probably should, then there are grounds for carrying out a control study. Note, however, that the control study stands by itself, that no commitment is required from either party to accept a causal explanation, and that there would have been no grounds to carry out the control study if both parties were in agreement that the spontaneous remission rate is appreciably lower than 80%. When no party would be willing to put its stake on the control condition, the use of controls would be as meaningless as is the testing of a null hypothesis that nobody believes to be true.

This shift of emphasis – away from differences between experimental and control conditions and towards differential predictions of either experimental or control results – is especially consequential for applied research that evaluates the effects of a treatment, an educational or a social program, an economic policy, and the like. In applied settings, it is always difficult and often impossible to both experiment and control. Classical methodology would lead to the conclusion therefore that rigorous research is impossible. A perfectly rational and scientific research script for these settings, however, is the following: Two or more parties articulate their predictions with respect to both the experimental and the control event, for example, revaluating versus not revaluating the dollar. The chances are that these predictions will diverge in both cases. So irrespective of which of the two materializes, it will be possible to evaluate a prediction.

One may object that for practical purposes the evaluation of predictions is not nearly as interesting as the evaluation of the treatments, programs, and policies themselves; in other words, that the betting model provides a solution to a different problem. However, this is a shortsighted argument. Evaluation of treatments, programs, and policies is also not a goal in itself; it is instrumental in improving decisions. Evaluation of predictions on the effect of decisions may be seen as another, more indirect way of improving decision-making. Public commitment to a prediction entails the risk of reputation loss, and will thereby exert a preventive influence. The focussing of a debate upon the empirical consequences of decisions will make the debate more to the point. Thus also from a practical point of view, the betting approach is not as powerless as it might seem.

Comparisons with other alternative approaches

Betting versus adversary statistics

A methodological model introduced by Woodworth (1976), named "adversary statistics", is the closest Bayesian counterpart of the betting model. The point of departure is the same: two parties each of whom entertains a prior probability distribution with respect to a certain population parameter θ. From there on, however, the two approaches diverge.

Under the adversarial statistical model, the aim of a party is to convince the other party through the accumulation of data. It is postulated that both parties will form their posterior beliefs in accordance with Bayes' theorem. Therefore, the gathering of relevant data will bring about the convergence of the posterior distributions. From the point of view of a particular party, the convergence will of course take place in the direction of that party's prior distribution. Through "pre-posterior analysis", a party can calculate the amount of data that will have to be gathered in order to reach a certain desired degree of convergence. The more the prior distributions are apart (reflecting the initial degree of dissensus), and the more peaked the adversary's prior is (reflecting the adversary's subjective certainty), the larger the amount of data that is needed.

It should be noted that the spirit of this model is not truly adversarial. In an antagonistic situation, both parties would be tempted to exaggerate their own subjective certainty, since their posterior distributions would be closer to their priors to the extent to which these priors are more peaked. In the limiting case in which a party would profess absolute prior certainty, no finite amount of data would be able to change its prior belief. Thus the adversarial statistical model cannot cope with cases in which the adversary has a strong investment in his or her belief, as in social or scientific prejudice. On the contrary, it would tempt people to arm themselves against empirical findings.

The consequences of the betting model are radically different. The more subjectively certain an opponent is, and the larger the difference of opinion, the *fewer* observations are needed to expect a particular gain in reputation. Furthermore the properness of the decision rule makes it irrational to profess exaggerated subjective certainty. In these respects, the betting reconstruction should be preferred over the adversary statistics model from a research policy point of view.

The scoring rule model

Van Naerssen (1961) has proposed a methodological model in which no restrictions are placed upon the predictive probability distribution of an investigator, thus admitting prior distributions instead of merely point hypotheses; once the data are observed, the prediction is scored with a proper scoring rule. This model differs from the betting model in that the investigator, as in classical hypothesis testing, plays against nature instead of an opponent.

In this model, the subjectively expected reputation is higher as the predictive probability distribution is more peaked. An investigator who happened to be certain that the sun will rise tomorrow would be gravely tempted to broadcast this prediction each day, since he or she would be certain to obtain a maximum score each time. More generally, the model would encourage safe bets on trivial issues. Note that the problem is not that the investigator would be tempted to exaggerate his or her subjective certainty: That manoeuvre would only lower one's subjectively expected reputation, since the scoring rule is proper. The problem is that the selection of issues would be conservative. Thus the model is unfit from a policy point of view.

The temptation to select safe issues is completely absent in the betting model. Quite to the contrary, since a prediction that is safe to one person tends to be shared by others, the predictive distributions would be practically identical, and the expected reputation gain would be practically zero. The betting model elicits daring predictions, albeit under the condition that the investigator is rationally convinced that his or her deviant position is tenable.

Statistical decision theory

A model in which two point hypotheses are tested against each other has been proposed by Koele (1979) in the framework of statistical decision theory. The model consists of calculating a sample size such that the two predictive distributions conditional upon the two hypotheses overlap less than 2α (e.g., $\alpha = 0.01$); it is decided that hypothesis A wins if the probability of observing the actual data X under hypothesis A, $p(X|A)$, is $> \alpha$, in which case $p(X|B)$ is automatically $< \alpha$, and vice versa.

This model would invite a participant who actually entertains a particular point hypothesis to submit a hypothesis very close to the opponent's, since that would enhance that participant's subjectively expected reputation (Hofstee and Nevels 1980). In other words, the scoring rule appears to be improper. Moreover, the model can only accommodate point hypotheses. In both respects the betting model is superior. I am in full agreement, however, with Koele's arguments against single null hypothesis testing.

Criticisms of the betting model

Strategic vulnerability

A valid criticism of the betting model, which will come as no surprise to decision theorists, is that the scoring rule is only proper when all parties try to maximize their subjectively expected reputations (Houtkooper 1981; Hofstee and Nevels 1981).

Consider the case in which the personal predictive probability of party A that a certain event will occur is 0.9 and B's probability is 0.5. According to eq. (5), A wins 0.24 units from B if the event occurs and loses 0.56 if it does not. The subjectively expected reputation of A is positive: $(0.9 \times 0.24) - (0.1 \times 0.56) = 0.16$. Assume, however, that A wishes to restrict her maximum loss. By submitting a probability of 0.7, for example, A loses only 0.24 units if the event does not occur, and wins 0.16 if it does. Note that the subjectively expected reputation of A is still positive: $(0.9 \times 0.16) - (0.1 \times 0.24) = 0.12$ (A's private probabilities, not the submitted probabilities, should of course be used in calculating A's subjectively expected reputation).

The real problem is not in the fact that A chooses to adopt a conservative strategy. If there were no implications for B's behavior, then A's strategy would not represent a serious threat to the betting system. However, the situation is not so simple.

Assume that B still wishes to maximize his subjectively expected reputation, true to the model. Suppose, however, that B suspects that A will not submit her private probability, but will submit a probability $r_A = \frac{1}{2}(p_A + r_B)$, where p stands for the private probabilities and r for the probabilities that are submitted for the record. That is, B suspects that as soon as r_B has been submitted, A will submit an r_A which is halfway between p_A and r_B. Suppose that p_A is known to B, as may be the case when the parties have been discussing the issue. It can then easily be shown that B should submit $r_B = \frac{1}{3}p_A + \frac{2}{3}p_B$ (cf. Hofstee and Nevels 1981). Since $r_B \neq p_B$, the decision rule is no longer proper with respect to B. In other words, as soon as the suspicion arises that a participant will depart from a reputation-maximizing strategy, the game is spoiled for the others. It should be added that these suspicions may arise quite naturally, since it is plausible that parties will not generally adhere to reputation maximization. Thus far, no satisfactory solution has been found for the vulnerability of the betting model in this respect. It is doubtful whether a solution can be envisaged at all; maybe no policy can ever be successful without an ultimate appeal to the fairness of the participants.

Other objections

An obvious objection against the betting model is that it is atheoretical in the sense that it does not reward theories, only predictions (e.g., Zwanenburg 1981). Within the framework of empirical science, this objection is easily dismissed and even reversed. The prime purpose of theories is to improve predictions; thus to the extent that theories are conducive to greater predictive accuracy, the construction of theories is rewarded and elicited by the betting model. On the other hand, theories that serve no predictive purpose but are motivated by aesthetic or metaphysical needs are not encouraged. It would thus seem that empirical science is better served by the model than by an indiscriminate plea for more theory. It is true that theories are not rewarded directly and as such; the reason is that other factors may have been the cause of a successful prediction. But investigators who feel that theory helps will enhance their subjectively expected reputations by engaging in theory construction. It is also true that impatient temperaments would seek short-term successes, but nothing in the model would prevent others from making a long-term investment in the form of theory development.

Another criticism, of a philosophical nature, is that theoretical points of view tend to be incommensurable (Zwanenburg 1981), so that any operationalization that is proposed by one party will be rejected by the other; consequently, bets between competing paradigms cannot take place. Several arguments can be brought against this criticism. In the first place, the betting scenario is an efficient way to decide whether two theories are in fact incommensurable; it precludes any one-sided presumptions that empirical findings should convince the other party. In the second place, the betting model does not encourage incommensurability claims: no bets, no reputation. This is as it should be. Undoubtedly, the history of science can be written in terms of fashions that have no common empirical ground, but from a normative point of view that is not a very satisfactory state of affairs. The betting model does not indulge in such relativism but counteracts it. In the third place, the incommensurability issue may well have been greatly exaggerated; it applies to grand theories that have little empirical content, but not to everyday practice in empirical science. A measure of the empirical content of a theory is the extent to which that theory denies predictive statements that are found plausible by others. In scientific practice, theories tend to have empirical content. The betting model favors such theories.

References

Campbell, D.T., 1969. 'Prospective: artifact and control'. In: R. Rosenthal and R.L. Rosnow (eds.), Artifact in behavioral research. New York: Academic Press.

Cohen, J., 1962. The statistical power of abnormal-social psychological research: a review. Journal of Abnormal and Social Psychology 65, 145–153.

De Finetti, B., 1965. Methods for discriminating levels of partial knowledge concerning a test item. British Journal of Mathematical and Statistical Psychology 18, 87–123.

Hofstee, W.K.B. 1977. De weddenschap als methodologisch model. Nederlands Tijdschrift voor de Psychologie 32, 203–217.

Hofstee, W.K.B. and K. Nevels, 1980. Scoring rules for competing predictions. Kwantitatieve Methoden 1(1), 75–90.

Hofstee, W.K.B. and K. Nevels, 1981. Do not take the betting model literally. Kwantitatieve Methoden 2(3), 70–72.

Houtkooper, J.M., 1981. The betting model: three competing predictions and the corrupting influence of bookmakers. Kwantitatieve Methoden 2(3), 65–69.

Koele, P., 1979. Afscheid van de nulhypothese. Kennis en Methode 3, 446–458.

Morrison, D.E. and R.E. Henkel (eds.), 1970. The significance test controversy. London: Butterworths.

Naerssen, R.F., van, 1961. A scale for the measurement of subjective probability. Acta Psychologica 17, 159–166.

Nevels, K., 1980. 'Appendix'. In: W.K.B. Hofstee, De empirische discussie: theorie van het sociaal-wetenschappelijk onderzoek. Meppel: Boom.

Pitz, G.F., 1978. Hypothesis testing and the comparison of imprecise hypotheses. Psychological Bulletin 85, 794–809.

Shuford, E.H., A. Albert and H.E. Massengill, 1966. Admissible probability measurement procedures. Psychometrika 31, 125–145.

Woodworth, G.G., 1976. t for two, or preposterior analysis for two decision makers' interval estimates for the mean. The American Statistician 30, 168–171.

Zwanenburg, M.A., 1981. Wetenschap als weddenschap: verlies door winst. Kennis en Methode 5, 261–272.