EDITORIAL

Can We Study Intelligence Using the Experimental Method?

HANS J. EYSENCK
Institute of Psychiatry
University of London

Cronbach (1957), at a famous presidential address to the American Psychological Association, argued that there were two disciplines in scientific psychology that ought to work together, but failed to do so: the experimental and the correlational. I have always argued, along the same lines, that correlational studies were concerned with taxonomy, but that causal arguments required an experimental approach (Eysenck, 1984). These admonitions have had little effect; traditional experimentalists still spurn the application of correlation-related concepts (intelligence, personality) to their studies, whereas correlationists make no effort to make their studies assume a more causal approach by using the experimental method. Boring (1951) put up the traditional case: "Is a test an experiment? And does the history of mental testing belong in the history of experimental psychology? Not really, although there is no sharp logical line of demarcation" (p. 570). According to Boring, the experimental approach uses an independent variable and determines what function the variation of a dependent variable is of the variation of the independent variable. In the field of testing, the primary variable is a difference of persons, which occurs at random and is not an independent variable, nor can it usually be changed experimentally.

Actually, the position of experimental psychologists is much less secure than at first appearance. Physicists can eliminate experimentally all factors other than the independent variable; psychologists cannot. Whatever the experiment, it involves persons, with different IQs, different personalities, different motivations, different attitudes, different emotions. As Gregory Kimble once told me (personal communication, April 7, 1971): "Ninety percent of the outcome of a conditioning experiment is determined before the subject is strapped in." That is, the subject's attitude, knowledge, treatment by the receptionist, and emotions determine his or her reactions over and above the independent variable as manipulated by the experimenter. As an example, consider the experimental studies of Shigehisa and Symons (1978) and Shigehisa, Shigehisa, and Symons (1973) on the
Urbantschisch (1883) hypothesis. Urbantschitsch postulated the existence of intersensory effects, suggesting that the perception of different modalities could be facilitated by simultaneous heteromodal stimulation. Dozens of experimental studies gave contradictory results. Shigehisa and his colleagues argued that these studies neglected the effect of transmarginal inhibition and its dependence on personality type. They succeeded in showing that, for tests of extraversion–introversion, results depended crucially on these personality differences. Here, as in many other senses, experimental results cannot be interpreted without recourse to personality and ability variables (Eysenck, 1984).

Conversely, the taxonomic studies carried out by correlational psychologists cannot provide any final resolution of this problem, because the favored polythetic methods lack any causal element that would decide between alternative theories. Such causal elements are obviously needed in view of the widespread disagreements in the fields of personality (Eysenck, 1991), or the long-continued argument about g (Horn, 1985). Causal attributions are of course frequently made, but they lack the proof that only experiment can provide. Boring (1951) was completely right in denying an absolute difference, with the line of demarcation not very sharp. Nature can provide changes in the independent variable that can be used in an experimental fashion; so can society. Twin studies are examples of the first; adoption studies of the second possibility. Astronomers, similarly, make use of changes in the independent variable produced naturally, in lieu of being able to do so themselves. Thus, Eddington used the 1919 eclipse of the sun to study the bending of a star’s light rays by the sun to test Einstein’s prediction, and astronomers use the regular pulsations of Cepheids, coupled with similarity of brightness, to measure stellar distances. There is a continuum from pure experimentation to pure correlation, but it is possible to discriminate on the basis of intervention. A study is experimental when an intervention occurs to alter the status of the independent variable, and changes in the dependent variable are noted. I want to argue for the usefulness of such experimental intervention in solving certain problems that arise in the interpretation of correlational data.

Let us look at the concepts of fluid (g_{f}) and crystallized (g_{c}) ability, introduced by Cattell (1963). Brody (1992) gave a good discussion of the argument. Can these notions be divorced from the fact that tests of g_{f} are usually timed and that tests of g_{c}, such as vocabulary, are usually untimed? The nature of these factors, derived from factor analytic analyses, depends on interpretation, inevitably subjective in part, and does not give us demonstrable conclusions. Can we use the Wechsler verbal and nonverbal scales as good measures of g_{c} and g_{f}? We can argue for and against, but can we experimentally demonstrate our conclusion?

As Brody (1992) emphasized, Cattell’s (1963) theoretical interpretation of g_{f} and g_{c} has several deductive implications; he suggested that changes in the biological state of the organism influence g_{f} more than g_{c}, and he mentioned nutrition as a possible variable. Recent work has used vitamin and mineral supplementation to test the hypothesis that such supplementation would raise the
nonverbal IQ of children on the Wechsler Intelligence Scale for Children (WISC), as a measure of $g_p$, but fail to do so for verbal IQ, as a measure of $g_c$. This prediction is based on the hypothesis that $g_f$ is the fundamental variable, whereas $g_c$ represents the knowledge acquired, differentially, by children of varying $g_f$. This agrees well with the results of confirmatory factor analysis and multidimensional scaling, locating $g_f$ tests at the center of the scaling diagram and at the top of the factorial dendrogram. We thus have an experimental paradigm to test hypotheses derived from correlational studies. The results of 10 studies have been reported elsewhere; I shall here only deal with major predictions and the final outcome (Eysenck & Schoenthaler, in press). Table 1 gives some details.

The major finding was that in all 10 studies surveyed, none found a significantly increased verbal IQ in the supplementation group, as compared with the placebo control group. In all 10 studies, the supplementation group did better

### Table 1

<table>
<thead>
<tr>
<th>Author</th>
<th>Ages (years)</th>
<th>Control Group N</th>
<th>Active Group N</th>
<th>Control Group Change</th>
<th>Active Group Change</th>
<th>p</th>
<th>Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bemona Schoenthaler</td>
<td>13-16</td>
<td>30</td>
<td>30</td>
<td>1.8</td>
<td>9.0</td>
<td>.01</td>
<td>+7.2</td>
</tr>
<tr>
<td>Nelsonh</td>
<td>7-12</td>
<td>105</td>
<td>105</td>
<td>-1.0</td>
<td>5.0</td>
<td>.05</td>
<td>+6.0</td>
</tr>
<tr>
<td>Crombic</td>
<td>11-13</td>
<td>44</td>
<td>42</td>
<td>1.5</td>
<td>3.9</td>
<td>.22</td>
<td>+2.4</td>
</tr>
<tr>
<td>Benton</td>
<td>13</td>
<td>87</td>
<td>80</td>
<td>2.0</td>
<td>5.0</td>
<td>.02</td>
<td>+3.0</td>
</tr>
<tr>
<td>Schoenthalerc</td>
<td>12-16</td>
<td>100</td>
<td>105</td>
<td>8.9</td>
<td>10.1</td>
<td>.01</td>
<td>+1.0</td>
</tr>
<tr>
<td>Nidich</td>
<td>8-9</td>
<td>16</td>
<td>18</td>
<td>4.9</td>
<td>9.8</td>
<td>.04</td>
<td>+4.9</td>
</tr>
<tr>
<td>Schoenthalerd</td>
<td>12-16</td>
<td>25</td>
<td>24</td>
<td>7.0</td>
<td>8.1</td>
<td>.27</td>
<td>+1.1</td>
</tr>
<tr>
<td>Schoenthalerc</td>
<td>18-21</td>
<td>90</td>
<td>96</td>
<td>6.3</td>
<td>7.1</td>
<td>.12</td>
<td>+0.8</td>
</tr>
<tr>
<td>Mσ</td>
<td>556</td>
<td>883</td>
<td>6.16</td>
<td>9.68</td>
<td>.001</td>
<td>+3.5</td>
<td></td>
</tr>
</tbody>
</table>

Note. From Eysenck and Schoenthaler (in press).

*Thirty controls received placebos and 26 received no pills.  
*The main index of nonverbal IQ (i.e., the WISC-R Coding Scale) is reported, but it is not a reliable measure of IQ.  
*Three strengths of supplement were compared with placebo. As strength increased, net gain increased with the exception of the strongest formula, but the difference between the two stronger formulas was not significant and probably chance.  
*Three strengths of supplement were compared with placebo. As strengths increased, net gain increased.  
*Two strengths of supplement were compared with placebo. The stronger formula produced the greater gains.  
*The probability of 10 independent trials each producing greater gains in the supplement group on the primary measure of nonverbal IQ is 0.5 to the 10th power, $p < .001$. 


than the placebo group on the nonverbal tests. This superiority was not always statistically significant; this failure was mainly due to such factors as insufficient statistical power (too small groups), too short periods of supplementation, or too small a supplementation of the recommended daily allowance (RDA). The mean IQ gain (supplementation over placebo) was 3.5 IQ points, but of course it had been predicted that the improvement would only occur in children with defective diets, that is, children with vitamin and mineral deficiencies. If only one in three children have such deficiencies, the overall level of 3.5 gives a misleading impression; for any two children with no improvement, there would be one with $3.5 \times 3 = 10.5$ nonverbal IQ points improvement. Note that all these studies were done on children who were considered to be adequately fed, and none on deprived inner-city children, nor children in third-world countries. None would have been considered poorly nourished.

In our own studies, we used blood analyses to determine nutritional status at the beginning and the end of the experiment. We used four groups of approximately 100 children, allocated randomly to a placebo group, a 50% RDA group, a 100% RDA group, and a 200% RDA group. We defined children who failed to increase their nonverbal scores after supplementation as nonresponders and argued that responders, defined as improving by twice the standard error of the estimate (i.e., 9 points more than the test–retest gain of 9 points) would have significantly lower vitamin and mineral levels than nonresponders. This we found to be so ($p < .01$) for vitamins, but not for minerals.

Clearly, vitamins are more important than minerals in producing an increase in $g_f$. Not all children showed an increase in vitamin levels after supplementation, even when their preintervention levels had been low; reasons for such failure of benefiting from supplementation are not clear.

Our prediction of nil influence of micronutritional supplementation on $g_c$ seems to be borne out by the 10 studies listed in Table 1 but is contradicted by two facts. Part of the Schoenthaler, Amos, Eysenck, Peritz, and Yudkin (1991) study was a routine educational test measuring knowledge acquired during the period of the supplementation; some of these tests showed significant improvement of the supplement groups over the placebo groups. But of course, $g_c$ tests normally test knowledge acquired during the course of the child’s life, and that would not be affected by subsequent supplementation. But the knowledge in these educational tests was acquired during the supplementation period, and hence any increase in $g_f$ would help the children so affected to improve their learning ability and show a (mild) improvement over placebo children.

In the other study that might seem to contradict our hypothesis, Schoenthaler, Doraz, and Wakefield (1986a) studied the results of dietary improvements in the food supplied to public school students in New York schools, revisions taking place during the 1979–1980, 1980–1981, and 1982–1983 academic years, no changes being made during 1981–1982. The diet changes consisted in the gradual elimination of synthetic colors, synthetic flavors, and selected preservatives.
At the same time, high-sucrose foods were gradually eliminated. This would correspond (very roughly!) to an improved vitamin and mineral supply, effected through dietary changes rather than through micronutrient supplementation.

The California Achievement Test was administered each year, and from it the national percentile rank of the school was determined. Results were very clear cut and are shown in Figure 1. From a position of 39th percentile in 1978–1979, there was a jump to the 47th percentile in the first year of the new diet, and from 47th to 51st when additional improvements were made. During the next year, there was no dietary change, and no further improvement. Finally, during 1982–1983, there was another dietary improvement, and another jump to the 55th percentile. There clearly was a remarkable improvement, due to the dietary changes imposed; Schoenthaler, Doraz and Wakefield (1986b) have discussed various alternative explanations, all of which could be eliminated.

These results fit in perfectly with the theory. Although scholastic improvement is measured by tests related to \( g_c \), the improvement would be conceptualized as due to increases in \( g_f \) mediating better learning ability resulting in scholastic achievement of a higher order. These experimental results lend strong support to those like Gustafsson (1984) and Snow, Kyllonen, and Marshalek (1984) who assigned a central role to \( g_f \) and a supporting role to \( g_c \), regarding it as an achievement variable mediating effects of \( g_f \). This experimental support should help in silencing alternative views (e.g., Gardner, 1983; Guilford &

![Figure 1. Effects of dietary changes on the standing of New York schools (from Schoenthaler, Doraz, & Wakefield, 1986a).](image-url)
Hoepfner, 1971; Horn, 1985). Other findings from these studies also support the theory. Thus, the effects of micronutrient supplementation are strongest with the youngest groups, weakest with the oldest. This is what one would expect, with effects being strongest at age levels where brain tissue is least strongly formed.

Another example of the possibility and usefulness of the experimental method in interaction with the correlational method can be found in the study of creativity (Eysenck, 1993, 1994a,b, 1995). Glover, Ronning, and Reynolds (1989) voiced their dissatisfaction with what they regarded as a “degenerating” research program in their *Handbook of Creativity*, and this dissatisfaction seems to stem largely from the purely correlational nature of the studies surveyed. As creativity is in so many ways an important aspect of intelligence and cognitive functioning, I tried to develop a model, the major features of which could be experimentally tested.

To begin with, I shall sketch the general background against which the theory in question was developed:

1. Creativity is not identical with g.
2. Creativity can be measured by tests of divergent thinking, unusual word association responses, preference judgments for complex visual stimuli, and so forth.
3. Such tests successfully predict creativity as shown in artistic and scientific achievements.
4. Tests of convergent thinking do not predict such creative achievements.
5. Associationist theories assert that creative achievement in tests and in real life is associated with a shallow associative gradient, and that rigid noncreative thinking is associated with a steep associative gradient.

There is good evidence for these brief summary statements (Eysenck, 1995). Several important further statements seem justified:

6. Creative people are often characterized, as far as personality is concerned, by a high degree of psychopathology.
7. Psychopathology is very frequently manifested by high achievers (geniuses) in the arts and sciences.
8. This psychopathology does not reach the level of actual functional psychosis.
9. This psychopathology is usually accompanied by a high degree of ego strength, normally negatively correlated with psychopathology.

From these facts, I deduced the following:

10. The personality dimensions of psychoticism (P) should be causally linked with creativity (Eysenck, 1992; Eysenck & Eysenck, 1976).
Psychoticism is conceived as a dimension of personality underlying functional psychotic disorders and predisposing high P scores to the development of such disorders given sufficient environmental stress.

The empirical literature has shown that in fact the connection between P and creativity, measured along the three lines mentioned under the second point in the preceding list, is quite strong (e.g., Eysenck, 1994b; Eysenck & Eysenck, 1976; Woody & Claridge, 1977) and even extends to high creative achievement (Götz & Götz, 1979a, 1979b). Given the associationist theory briefly mentioned under Point 5, it would follow that high P scorers (and functional psychotics) have a shallow associative gradient. It is this hypothesis that has been exposed to experimental testing (Eysenck, 1995). We are concerned with a theory that would encompass the cognitive peculiarities of schizophrenics, the creativity of the genius, and the experimental studies of associative processes. It may be useful to start with a well-established theory, namely Cameron’s notion of “overinclusion” (Cameron, 1947; Cameron & Magaret, 1950). Cameron believes that schizophrenics’ concepts are over-generalized. Schizophrenics are unable to maintain the normal conceptual boundaries and incorporate into their concepts elements, some of them personal, which are merely associated with the concept, but are not an essential part of it. Cameron used the term overinclusion to describe this abnormality and reported that in working on the Vigotsky test, and a sentence completion test, schizophrenics were unable to preserve the “conceptual boundaries” of the test. In solving a problem the schizophrenic “included such a variety of categories at one time, that the specific problems became too extensive and too complex for a solution to be reached” (Cameron, 1939, p. 263). A fair number of experiments have been carried out to investigate this theory. These have been reviewed elsewhere (Payne, 1960; Payne, Matusek, & George, 1959).

Payne et al. (1959) have suggested that it is possible to reformulate Cameron’s (1939) theory of overinclusion in a slightly more general way so that a number of predictions follow from it. Concept formation can be regarded as largely the result of discrimination learning. When a child first hears a word in a certain context, the word is associated with the entire situation (stimulus compound). As the word is heard again and again, only certain aspects of the stimulus compound are reinforced. Gradually, the extraneous elements cease to evoke the response (the word), having become “inhibited” through lack of “reinforcement.” The inhibition is in some sense an active process, as it suppresses a response which was formerly evoked by the stimulus. Overinclusive thinking may be the result of a disorder (failure) of the process whereby inhibition is built up to circumscribe and define the learned response (the word or “concept”). In short, it could be an extreme degree of “stimulus generalization.”

The same theory can be expressed in different terms. All purposeful behavior depends for its success on the fact that some stimuli are “attended to” and some other stimuli are ignored. It is a well-known fact that when concentrating on one task, normal people are quite unaware of most stimuli irrelevant to the task. It is
as if some "filter mechanism" cuts out or inhibits the stimuli, both internal and external, which are irrelevant to the task in hand, to allow the most efficient "processing" of incoming information. Overinclusive thinking might be only one aspect of a general breakdown of this filter mechanism.

A similar concept to overinclusion is that of "allusive" thinking, characteristic of many schizophrenics on object-sorting tests. McConaghy and Clancy (1968) demonstrated that this type of thinking existed widely in less exaggerated forms in the normal population, showed similar familial transmission in schizophrenics and nonschizophrenics, and was akin to creative thinking. Dykes and McGhie (1976) actually demonstrated that highly creative normals scored as highly allusive on the Lovibond object-sorting test, as did schizophrenics. The low creative normals tended to produce conventional, unoriginal sortings, whereas the highly creative normals and the schizophrenics tended to give an equal proportion of unusual sortings. "This supports strongly that a common thinking style may lead to a controlled usefulness in normals and an uncontrollable impairment in schizophrenics" (Woody & Claridge, 1977, p. 243).

What measurable agency can cause this cognitive inhibition which in normal people prevents loose, allusive overinclusiveness, and fails to do so in schizophrenics? A major candidate for the job is latent inhibition (Lubow, 1989).

Latent inhibition is defined by an experimental paradigm which requires, as a minimum, a two-stage procedure. The first stage involves stimulus pre-exposure, that is, the to-be-conditioned stimulus (CS) is exhibited without being followed by any unconditioned stimulus (UCS); this leads theoretically to the CS acquiring a negative salience, that is, it signals a lack of consequences and, thus, acquires inhibitory properties. The second stage is one of acquisition, that is, the CS is now followed by an UCS and acquires the property of initiating the unconditioned response (UCR). Latent inhibition (LI) is shown by increasing difficulties of acquiring this property, as compared with lack of pre-exposure. With humans, there is a masking task during pre-exposure to the CS. For instance, the masking task might be the oral presentation of a series of syllable pairs, whereas the CS would be a white noise randomly superimposed on the syllable reproduction. The control group would be exposed only to the syllable pairs, without the white noise. In the test phase, the white noise is reinforced, and participants are given scores according to how soon they discover the rule linking CS with reinforcement. LI would be indicated by the group having the preexposure of the white noise discovering the rule later than the control group.

There are more complex, three-stage procedures, but these complications are not crucial to our argument (Lubow, 1989). Clearly, the latent inhibition phenomenon involves cognitive elements; these are emphasized by Lubow (1989) in terms of his conditioned attention theory. According to this theory, nonreinforced preexposure to a stimulus retards subsequent conditioning to that stimulus because during such pre-exposure, the participant learns not to attend to it. The theory is based on the use of attention as a hypothetical construct, with the prop-
properties of a Pavlovian response, and on the specification of reinforcement conditions that modify attention.

Does latent inhibition correlate negatively, as it should, with P? Several studies have shown that it does; high P scorers show little if any latent inhibition, whereas low P scorers show considerable latent inhibition. We may go even further. Theories of schizophrenia prominently involve dopamine-D2 (Gray, Felton, Rawlins, Hemsley, & Smith, 1991); does P show a similar correlation? Gray, Pickering, and Gray (1994) have indeed found a substantial relationship of this kind, suggesting a causal sequence, rather like this:

\[
\text{Schizophrenia} 
\text{DNA} \rightarrow \text{Dopamine D2} \rightarrow (\text{lack of}) LI \rightarrow P 
\downarrow 
\text{Creativity}
\]

The alternatives implied by the two arrows at the right may be decided by the amount of stress experienced, quantitative differences in dopamine activity, and/or the presence of such positive personality factors as high ego strength (Barron, 1968, 1969). The latent inhibition variable can of course be manipulated by the use of agonistic and antagonistic drugs, and work along these lines has given clear-cut results in line with prediction (Eysenck, 1995). Latent inhibition seems a good, solid experimental basis for the notions of a shallow associative gradient, overinclusiveness, or allusive thinking, and thus for creativity.

It will be clear why I feel that the study of one important aspect of intelligence—creativity—can be furthered dramatically by the introduction of causal theories and their experimental study. Purely correlational investigations are a useful beginning, but they require extension into the field of general psychological laws and experimental study, as Spearman (1923, 1927) recognized explicitly many years ago. The tragedy of psychology as a scientific discipline is that even nowadays the gulf between correlational and experimental approaches that Cronbach (1957) complained about is still as wide as ever and shows little sign of being bridged. Only by combining these two disciplines within psychology, can psychology acquire a truly scientific stature. The examples given here will serve to illustrate the possibilities held out by this approach. The recent explosion of experimental studies into the speed of mental processes and the psychophysiological bases of cognitive behavior (Barrett & Eysenck, 1992; Eysenck, 1986a, 1986b) promises to lead to a more theory-oriented and experimental approach to the whole problem of intelligence.

REFERENCES


Urbantschitsch, V. (1883). Über den einfluss von Trigeminusreizen auf die sinnesempfindungen
insbesondere auf den gesichtsinn. [On the influence of Trigeminus stimulation on sensory receptors, particularly visual receptors]. Archive fuer die gesamte Physiologie, 30, 129–175.