



HOKKAIDO UNIVERSITY

Title	MIND AS A CULTURAL ACHIEVEMENT : IMPLICATIONS FOR IQ TESTING
Author(s)	COLE, Michael; コール, マイケル
Description	
Citation	乳幼児発達臨床センター年報, 3, 7-22
Issue Date	1981-03
Doc URL	https://hdl.handle.net/2115/25189
Type	departmental bulletin paper
File Information	3_P7-22.pdf



MIND AS A CULTURAL ACHIEVEMENT : IMPLICATIONS FOR IQ TESTING

Michael Cole

University of California, San Diego

For almost as long as there have been IQ tests, there have been psychologists who believe that it is possible to construct "culture free" tests. The desire for such tests springs directly out of the purposes for which tests of general intellectual ability were constructed in the first place : to provide a valid, objective, and socially unbiased measure of individual ability. Our society, founded upon the principle that all men are created equal, has never lived easily with the recognition of enormous *de facto* social inequality. We need a rationale for such inequality and our traditions strongly bias us to seek the causes of inequality in properties of the individual, not society. At the same time, we realize that social and economic inequality in can be the causes of individual intellectual inequalities, as well as their consequences.

What would be more ideal, then, than a psychological test that could measure intellectual potential that is based equally on the experience of people from all cultures. Can't we find universals in human experience and construct a test on this basis? Some psychologists have claimed not only that such tests are possible principle, but have been applied in practice (e. g., Jensen, 1979).

In this paper I will defend the proposition that the notion of a culture *free* IQ test is an absurdity. By virtue of a common phylogenetic-heritage, newborn children the world over share a common readiness to fill *homo sapien's* ever changing niche on the planet Earth. But all higher psychological processes are shaped by our experiences and these experiences are culturally organized.

19TH CENTURY PRECURSORS

Having made these assertions, I want to provide the evidence upon which they are based. My own personal strategy for thinking about these matters is to think my way back in time into the 19th century, when such devices came into existence. I have found it helpful to study the conditions in science and society that allowed some scholars to believe it is possible to assess mind independent of culturally organized experience. In this chapter I will review some of that history, emphasizing the logic of the enterprise. I am focusing on an *anthropological* perspective on testing, but it should become clear that

Editor's note : Dr. Michael Cole is Professor of Developmental Psychology, Laboratory of Comparative Human Cognition, Center for Human Information Processing, University of California, San Diego. During August and September, 1980, as a visiting professor of JSPS, Dr. Cole visited RCCCD where he gave a lecture and held a colloquium with the staff and graduate students.

anthropology and psychology have always been linked in shaping our understanding of the relation between experience and mind, even when this link is obscured by divergent methods and theories.

The several decades just preceding this century provide a useful starting point from which to trace theories of culture and cognitive development, because it was during this period that both anthropology and psychology took shape as disciplines. Before that time, say the 1860's, there was no distinctive body of methods for the study of the "human sciences" nor had scholars with different theories been institutionally separated into separate disciplines the way they are today. Obvious differences in technological achievement between peoples living in different parts of the world were common knowledge. Theorizing about sources of these differences had produced rather general acceptance of the notion that it would be possible to study the history of humanity by a study of contemporary peoples at different "levels of progress." E. B. Tylor summarizes, in what he calls "mythic fashion", the general course of culture that most of his fellow scholars would have adhered to :

We may fancy ourselves looking on Civilization, as in personal figure she traverses the world ; we see her lingering or resting by the way, and often deviating into paths that bring her toiling back to where she had passed by long ago ; but direct or devious, her path lies forward, and if now and then she tries a few backward steps, her walk soon falls into a helpless stumbling. It is not according to her nature, her feet were not made to plant uncertain steps behind her, for both in her forward view and in her onward gait she is of truly human type. [Tylor, 1958, p. 69]

Tylor's choice of imagery for "Civilization" nicely reveals another basic assumption which he and many of his colleagues made : there is no principled distinction between mind and society. The condition of culture among the various societies of mankind Tylor tells us, reveal basic information about the laws of human thought. He even adopted the notion of a "mental culture" which he expected to be high or low depending upon the other conditions of culture with which it was associated.

William Spencer, writing at about the same time, shared Tylor's belief in the fusion of mental and cultural phenomena. He also drew a very tight analogy between cultural development on the one hand and mental development on the other.

During early stages of human progress, the circumstances under which wandering families and small aggregations of families live, furnish experiences comparatively limited in their numbers and kinds ; and consequently there can be no considerable exercise of faculties which take cognizance of the *general truths* displayed throughout many special truths. (Spencer, 1886, p. 521)

Spencer invites us to consider the most extreme case ; suppose that only one experience were repeated over and over again, such that this single event comprised all of the person's experiences. In this case, as Spencer put it, "the power of representation is limited to reproduction of this experience" in the mind. There isn't anything else to think *about!* Next we can imagine that life consists of two experiences, thus allowing at least elementary comparison. Three experiences add to the elementary comparisons, and elementary generalizations that we make on the basis of our limited (3) experiences. We

can keep adding experience to our hypothetical culture until we arrive at the rich variety of experiences that characterizes our lives. It follows from this line of reasoning that generalizations, the "general truths" attainable by people, will be more numerous and more powerful, the greater one's experience. Since cultures provide experience, and some cultures (Spencer claimed) provide a greater diversity of experience than others, a neat bond between cultural progress and mental progress is cemented.

Although such evolutionary schemes seemed almost transparently obvious in the enthusiasm following publication of Darwin's *Origin of Species*, events toward the close of the 19th century proved that there could be a great deal of disagreement about the relation between culture and thought, despite the compelling story constructed by people like Tylor and Spencer. One set of disagreements arose when scholars started to examine more closely the data used to support conclusions about relations between cultures, especially claims for historical or evolutionary sequences. Quite a different set of arguments arose around conflicting claims about mental processes.

The seed of disagreements concerning cultural sequences can be found in Tylor's own work. The main criteria for judging the stage of a culture were the sophistication of industrial arts (including manufacturing techniques for metal tools, agricultural practices) and "the extent of scientific knowledge, the definitions of moral principles, the conditions of religious belief and ceremony, the degree of social and political organization, and so forth" (Tylor, 1958, p. 27). However, in Tylor's words, "If not only knowledge and art, but at the same time moral and political excellence, be taken into consideration" it becomes more difficult to scale societies from lower to higher stages of culture.

This latter theme in Tylor's work was taken up by Franz Boas, who submitted the cultural evolution position to a devastating critique at the close of the 19th century. On the basis of his own ethnographic work, Boas concluded that a great deal of the evidence apparently supportive of evolutionary schemes was so deeply flawed that no clear conclusions ranking one culture above another could be accepted (Boas, 1911). Boas did more than show the flaws in evolutionists' data and arguments concerning *culture*; he also delighted in showing that examples of "primitive mind" produced as part of this argument were based on misunderstandings.

Consider the following example from Boas' classic, *The Mind of Primitive Man*, which repeats evidence used by Spencer to make some generalizations about properties of primitive mind :

In his description of the natives of the west coast of Vancouver Island, Sproat says, "The native mind, to an educated man, seems generally to be asleep. . . . On his attention being fully aroused, he often shows much quickness in reply and ingenuity in argument. But a short conversation wearies him, particularly if questions are asked that require efforts of thought or memory on his part. The mind of the savage then appears to rock to and fro out of mere weakness."

Spencer's text goes on to cite a number of similar anecdotes corroborating this point. But Boas produces an anecdote of his own.

I happen to know through personal contact the tribes mentioned by Sproat. The questions put by the traveller seem mostly trifling to the Indian, and he naturally soon tires of a conversation carried on in a foreign language, and one in which he finds nothing to interest him. As a matter of fact, the interest of these natives can easily be raised to a high pitch, and I have often been the one who was wearied out first. Neither does the management of their intricate system of exchange prove mental inertness in matters which concern them. Without mnemonic aids to speak of, they plan the systematic distribution of their property in such a manner as to increase their wealth and social position. These plans require great foresight and constant application. (p. 128)

Thus, Boas tells us that the entire scheme was wrong. Cultures cannot be ranked in evolutionary age as a basis for comparison and "mind" cannot be seen as rank in developmental age.¹

Finally, and very importantly, Boas was a leader in a subtle, but essential change in anthropological thinking about the concept of culture itself. Educated in Germany, Boas had begun his career imbued with the Romantic concept of "Kulture," the expression of the highest attainments of human experience, as expressed in the arts, music, literature and science. This is the conception of culture that allowed Tylor to talk about "the conditions of culture among various societies ;" Tylor, like Boas as a young man, conceived of culture as something groups and individuals had more or less of. It was a singular noun : one talked of higher or lower *culture*, not more or fewer *cultures*. By the same route that led him to deny the basis for ranking cultures in terms of a hypothetical, evolutionary sequence, Boas arrived at the idea that different societies create different "designs for living," each representing a uniquely adapted fit between their past and their present circumstances in the world. This point of view is central to anthropology, and it clearly has to be taken into account if we want to rank the intellectual achievements (levels of mental development) of people growing up with different cultural experiences. It renders simple more/less comparisons of cultures difficult and restricted, with parallel affects on our inferences about mind.

ENTER PSYCHOLOGY

As we entered the 20th century, anthropology was still pursuing its goal of reconstructing the history of mankind by studying cultures in different parts of the world. But that goal was now blocked by serious methodological problems (such as those raised by Boas) that needed to be settled before further theoretical progress could be made.

Psychology's birth is usually dated back to 1879, when Wilhelm Wundt officially opened an experimental laboratory in Leipzig. The exact date isn't important, because several laboratories opened almost simultaneously in different industrialized countries. But the *reason* for these laboratory openings are very important indeed.

Boas' critique of developmental theories, whether of mind or culture, produced controversy in both domains of inquiry. Boas earned the enmity of anthropologists who believed his criticisms of their general theories unjust ; they sought to rescue the more general theories, criticizing Boas and his students of "historical particularism" (to use

1. Boas also demonstrates the total hopelessness of deducing cultural differences from any differences, real or imagined, in genetic makeup.

Harris' apt phrase). While new competitors for an overall approach to understand historical links between cultures became a central activity for the new discipline of anthropology, psychologists were people who took up the half of Boas' critique, problems of specifying mental mechanisms.

The major difficulty facing those who became psychologists was to devise methods for specifying pretty exactly what sorts of activity an individual engages in at those times we want to make claims that some sort of "thinking" is going on. No one could be very precise about what they meant when they referred to a mental process. Competing claims were evaluated by constructing settings to control as exactly as possible the kind of events a person experienced and to record the kinds of responses these experiences evoked. Since the presumed processes were not observable (they were, as we say, "psychological"), psychologists spent a great deal of time and ingenuity devising ways to pin down what these non-observable processes might be. The rapidly growing ability to control electricity and to build precision machinery was exploited to the fullest ; the early psychology laboratories were marvels of inventions. Their instruments allowed psychologists to present people carefully controlled lights and tones for carefully controlled intervals and to measure precisely the time it took to respond. In their search for ways to make mind observable, they used electrophysiological devices to record internal, organic functioning. The discipline of "psychophysics" advanced appreciably in its quest to related psychological phenomena of an elementary order (discriminating tones, judging hues). There were even hopes of uncovering a "cognitive algebra" by carefully comparing reaction times to stimuli of various complexities arranged to reveal steps in the thought process.

The activities of the psychologist and the anthropologist soon contrasted very dramatically. The psychologist brought people into the laboratory where behavior could be constrained, stimuli controlled, and mind made visible. The anthropologist wandered the world talking to people, observing their customary behavior, and seeking clues about the factors that made one design for living different from another.

Whereas the anthropologists continued to concentrate on gathering data that would permit firm statements about historical relations between cultures, scholars who came to identify themselves as psychologists concentrated on resolving arguments about thinking such as those illustrated in the passage quoted from Boas. Just as anthropology evolved careful field techniques to disambiguate competing claims about "culture", psychologists developed the laboratory experiment as a way to test competing claims about "mind".

There occurred, in effect, a division of labor in the "humane sciences," a division that was primarily a matter of scientific strategy in the beginning : progress required some concentrated work on specialized sub-topics. The overall task remained the same for everyone : how do human beings come to be the way they are?

ENTER TESTING

Despite an increasing gulf between scholars who called themselves psychologists and those who called themselves anthropologists, it was not long before these two areas of inquiry were brought together again. At the end of the 19th century Francis Galton, in England, set out to test hypotheses about mental differences among people using the newly devised psychological techniques. His concern was not differences between people

growing up in different cultures. Rather, he studied people growing up in different families. He sought the inherited sources of variability in mental abilities. Significantly, his tests were theoretically motivated ; he believed that speed of mental processing was central to intelligence so he created tests of rapid processing of elementary signals. Galton succeeded in finding differences among Englishmen on such tests as simple reaction time to a pure tone, but he did not succeed in relating these "psychological test" differences to human characteristics of greater interest to him such as scientific excellence or musical ability. Galton's tests, based on an oversimplified model of the human mind and the highly controlled procedures adopted from the laboratory appropriate to testing his theory, were not taken up by society. However, in creating an early precursor of existing IQ tests, Galton did begin the development of the statistical techniques that would be necessary to show how test differences co-relate with interesting behavioral differences. The difficulties that Galton encountered in trying to demonstrate that he was testing abilities of general significance were a direct stimulus to the development of that branch of applied mathematics known as statistics, upon which current testing technology relies so heavily. The fact that these difficulties have not been resolved, despite great progress in the technology for evaluating the theory, is a key problem that remains to be dealt with.

Galton did all of his work in England, but other Englishmen, including W. H. R. Rivers, travelled to the Torres Straights northeast of Australia, to see if psychological tests could be used to settle disputes over cultural differences in cognition. Rivers was in some senses an antique. He was both anthropologist and psychologist which meant that he considered both the evidence of his tests and evidence provided by observation of the people he went to study when he made statements about culture and thought. His conclusions were consistent with Galton's data on individual differences ; natives differed from each other on such simple tasks as their ability to detect a gap in a line, or their recognition of colors. But there were no impressive differences between the natives of the Torres Straights and Englishmen.

It would appear on the basis of this evidence that there are no cultural differences in thinking, at least no differences consistent with what we had been led to believe by Tylor, Spencer, and many others. However, it could be (and was) argued, that the *important* ways in which cultural differences cause mental differences weren't even tested by Rivers and his associates. After all, Galton, had found no relation between responses to his psychological tests and other presumed indicators of intelligence ; why would anyone, then, expect cultural differences? Perhaps the experiments, limited as they were to *elementary* psychological processes, simply failed to implicate *higher psychological processes* at all. What we needed were tests of higher psychological processes that could be used to compare people from different cultures or different people in the same culture.

This distinction between elementary and higher processes pinpoints a weakness in the basic foundations of experimental psychology, a weakness acknowledged by Wundt, its founder. It is impossible, Wundt believed, to study *higher* psychological functions in experiments because it is impossible to construct appropriately controlled environments of the needed complexity. Wundt believed that scientists should use *ethnological* evidence and folklore if they want to discover the properties of the mind that get constructed on the basis of the elementary processes that he studied in the laboratory.

Wundt's doubts about the experimental method have not been accepted in psychology, but they are very germane to understanding problems with cross-cultural developmental research, as we shall see. These doubts were not accepted because they put psychologists in a very difficult bind. Psychology had been founded on the principle that without carefully controlled environments, it is not legitimate to make statements about how the mind works. But a great many of the questions about how the mind works that interested psychologists and anthropologists alike clearly refer to "higher" psychological processes such as logical reasoning and inference. When Wundt gave up on the idea that such processes could be studied in the laboratory, he was, it seemed, robbing psychology of most of its interesting subject matter. For psychologists, the inability to study higher psychological processes in the laboratory meant that they could not be studied at all. Rejecting this conclusion, many psychologists were attracted to theories claiming that complex processes are compounded of simple ones. The basic task was to understand the elements, before tackling the compound. Relatively simple experimental models thrived, but complex behavior was rarely dealt with.

BINET'S STRATEGY.

The major push for research on more complex human problem solving came from a source seemingly outside the scientific community, although respected psychologists were involved. Early in this century, Alfred Binet was asked to deal with a practical, social problem. With the growth of public education in France, there was a growing problem of school failure, or at least, severe school underachievement. It seemed not only that some children learned more slowly than others, but that some children, who otherwise appeared perfectly normal, didn't seem to benefit much from instruction at all. Binet and his colleagues were asked to see if they could find a way to identify slow learning children at an early stage in their education. If such identification were possible, special education could be provided them, and the remaining children could be more efficiently taught.

The subsequent history of IQ testing has been described too frequently to bear repetition here, but a sketch of the basic strategy of research is necessary as background to understand just how deeply IQ tests are embedded in cultural experience.

To begin with, early test makers had to decide what to test for. The decision seemed straightforward. They wanted to test people's ability to perform the kinds of tasks that are required by schools. They observed classrooms, looked at textbooks, talked to teachers, and used their intuitions to arrive at some idea of the many different kinds of knowledge and skills that children are eventually expected to master in school.

What Binet and his colleagues found was not easy to describe briefly, as anyone who has looked into a classroom can quickly testify (and all of us have done so, or we would not be reading these words). There was a very obvious need to understand graphic symbols, such as alphabets and number systems. So recognition of these symbols was tested. But mastery of the rudiments of these symbols was not enough. Children were also expected to manipulate these symbols to store and retrieve vast amounts of information, to rearrange this information according to the demands of the moment, and to use the information to solve a great variety of problems that had never arisen before in the experience of the individual pupil. Thus, children's abilities to remember and carry out

sequences of movements to define words, to construct plausible event sequences from jumbled picture sequences, and to recognize the missing element in graphic designs, were tested (along with many other components of school-based problems).

It was also obvious that to master more and more esoteric applications of the basic knowledge contained in alpha-numeric writing systems, pupils had to learn to master their own behavior. They had not only to engage in a variety of "mental activities" directed at processing information ; they also had to gain control over their own attention, applying it not according to the whim of the moment, but according to the whim of the teacher and the demands of the text.

It was clearly impossible to arrive at a single sample of all the kinds of thinking required by "the" school. Not only was there too much going on in any one classroom to make this feasible ; it was equally clear that the school required different abilities from children of different ages. Binet realized that estimates of "basic aptitude" for this range of material would depend upon how much the child had learned about the specific content before she/he arrived at school, but he felt knowing a child's current abilities would be useful to teachers anyway.

In the face of these difficulties, Binet decided to construct a sample of school-like tasks appropriate for each year of education, starting with the elementary grades, and reaching into higher levels of the curriculum. He would have liked to sample so that all essential activities were included in his test and that tasks at one level of difficulty would be stepping stones to tasks at the next higher level. But because no firmly based theory of higher psychological functions existed, Binet had to rely on a combination of his own common sense and a logical analysis of tasks that different classrooms seem to require (for example, you have to be able to remember three random digits before you can remember four ; you have to know the alphabet before you can read). He also hit on the handy strategy of letting the children themselves tell him when an item selected for the test was appropriate. Beginning with a large set of possible test questions, Binet hunted for items that half the children at a given age level could solve. An "average" child would then be one who solved problems appropriate to his/her age level. Keeping items that discriminated between children of different ages (as well as items that seemed to sample the activities demanded of kids in their classrooms), he arrived, with help from his colleagues, at the first important prototype of the modern IQ test.

Of course a great deal of work has gone into the construction of tests since Binet's early efforts, but the underlying logic has remained pretty much the same : sample the kinds of activities demanded by the culture (in the form of the problems it requires that its children master in school) and compare children's performance to see how many of these activities they have mastered. Children who have mastered far less than we would expect given a comparable sample of kids their own age are those who will need extra help if they are to reach the level expected by the culture.

This strategy is perfectly reasonable, so long as we stay within the framework that generated the item selection procedures in the first place. However, much to the disapproval of Binet, people found new uses for these tests of school-based knowledge that carried with them the seeds of the current disputes over IQ testing. Although Binet specifically warned against the procedure, his test and tests like it began to be used as

measures of an overall aptitude for solving problems *in general*, rather than *samples* of problem solving ability and knowledge *in particular*. Those engaged in such extrapolations acknowledged that in principle it is important to make certain that everyone given the test has an equal opportunity to learn the material that the test demands. But in practice there was no way to guarantee this essential prerequisite for making comparative judgements about basic abilities.

These are important issues in thinking about applications of IQ testing, and they are extensively discussed in the psychological literature. However, it is not until we back up and examine the possible significance of Binet's work in the light of anthropological scholarship that we can see just how limited an enterprise IQ testing was at the beginning, and how restricted it remains today.

A THOUGHT EXPERIMENT IN TEST CONSTRUCTION

A good starting point for this re-examination is to think about what sort of activity Binet would have engaged in if he had been a member of a cultural group vastly different from his own. As a sort of "thought experiment" let's suppose that a "West African" Binet has taken an interest in the kinds of knowledge and skill that a child growing up in his part of the world would need to master as an adult. To make the thought experiment somewhat concrete, I will do my supposing about the tribal groups inhabiting the interior of Liberia, principally the Kpelle people among whom I have worked and about whom a good deal of relevant information is available (e. g., Bellman, 1975 ; Cole, Gay, Glick and Sharp, 1971 ; Gibbs, 1965).

Following in the footsteps of his French model, our Liberian Binet would want to make a catalogue of the kinds of activities that children are expected to master by their parents and the village elders. People in rural Liberia make their living by growing rice and other crops, which they supplement by meat and fish when these scarce commodities can be obtained. Rice farming is physically difficult work that demands considerable knowledge and planning for its success, but as practiced by the Kpelle, it is not a technologically sophisticated enterprise. It is carried out using simple tools such as a machete, to cut the underbrush, fire, to burn the dry brush, vines to tie together fence posts in order to keep out animals, and slingshots to harass (see Gay, 1973, for a much better account of this process). Other aspects of Kpelle material culture are also relatively simple, although in every case, the proper use of tools requires a good deal of knowledge about how the tools are supposed to be used. There is division of labor among Kpelle adults (men hunt, women do most of the fishing ; men cut the bush on the farms, women plant the seed, children guard the crops) but far more than is true of contemporary America, everyone pretty well knows what there is to know about adult economic activities. There are some specialists (blacksmiths, bonesetters, weavers) whose work is an exception to this generalization, and study of their activities would certainly be important.

Of course, there is more to getting through life as a Kpelle than growing rice or weaving cloth. All descriptions of the social organization of Kpelle life stress that as in America, knowledge of the social world is essential to adult status. Kpelle people are linked by a complex set of relations that control how much of the resources available to the society actually get to the individual.

Faced with this situation, how should our West African Binet proceed? Should he sample all the kinds of activities valued by adults? This strategy is almost certainly unrealistic. Even allowing for the possibility that aspects of technology make it reasonable to speak of the Kpelle as a "less complex" society than our own, it is very complex indeed. No anthropologist would claim to have achieved a really thorough description of even one such society. Moreover, like Tylor, he would have to admit the possibility that in some respects, Kpelle society provides members with more complex tasks than we are likely to face. Since it is unreasonable in Liberia, as it is in the United States, to think that we can come up with a test that samples *all* types of Kpelle adult activities, why not follow Binet's example and sample an important *subset* of those activities? From an anthropological perspective, schools are social institutions for assuring that adult knowledge of highly valued kinds gets transmitted to a society's next generation (it must be transmitted, or there would be no later generations!). While the school is not likely to be a random sample of life's tasks, it is certainly a convenient place to sample activities that adults consider important, activities that are complex enough to make it unlikely that kids would learn what they need to know simply by "hanging around."

So, our Liberian Binet might decide to search for some institutions in his society that correspond roughly with the basic goals of schooling in ours. Not all societies readily manifest such institutions, leading anthropologists to speak of socialization, as the broadest relevant category. Fortunately for discussion, in the case of Liberia, he would undoubtedly discover the existence of institutions called "bush schools" in the creole vernacular.

There are no detailed accounts of the curriculum of the bush school. The three or four years that youngsters spend are organized by town elders who are leaders in the secret societies that control a variety of esoteric information. This material cannot, on pain of death, be communicated to outsiders (see Bellman, 1975, for the most detailed account of these practices). However, we know enough about aspects of bush school activities to continue our hypothetical research; we know that youngsters learn to farm, construct houses, to track animals, shoot birds, and carry out a variety of adult economic activities (children live apart from their home villages in something like a scouting camp during their time in bush school). They are also instructed in the important lore of the group. This lore is communicated not only in a variety of ceremonies, but in stories, myths, and riddles. So, let us suppose that our West African Binet decided to use "successful execution of bush school activities" as the abilities he wanted to sample.

Again, like Binet, our researcher would not be able to sample *all* such activities for his test, nor would he want to. He would not, for example, want to sample activities that all children knew how to accomplish *before* they got to school, nor would he want to sample activities considered so universally accessible that everyone mastered them well before the end of schooling. This information wouldn't help him pick out those children who needed extra instruction. Instead, he would seek those activities that discriminated among children, activities that some mastered far earlier than others, and perhaps activities that some only mastered in later life. Once these Binet-like restrictions had been placed upon the activities selected for study, our hypothetical researcher could begin selecting tasks on which he could base test items.

In considering what sort of test would emerge, its useful first to consider what activities would be excluded as well as those included. Cutting brush or sowing rice seed probably wouldn't be the test ; everyone knows how to do that before he or she gets to school. Nor would anyone spend time explicitly teaching children common vocabulary. However, there would be explicit instruction in such tasks as constructing houses and identifying leaves that are useful in different kinds of medicine. There would also be some mechanism for insuring that the history of the group, its laws and customs, were taught to everyone often in the form of stories and dances. Finally, some children would be selected for specialist roles that would require special tests (bonesetter, weaver, midwife, blacksmith, hunter and so on), these children would receive additional instruction.

Looking at those areas where instruction might be considered important, we can see many candidate activities for testing. We might want to see if children had learned all of the important leaf names for making medicine. Riddles are often important parts of stories and arguments, so we could test to see how many riddles children know and how adept they are at interpreting them. The specialties would be a rich source of test material, especially if we thought that rational testing of ability to perform like adults would improve the quality of our cloth or machetes. In short, it seems possible in principle, to come up with test items that could perform functions in Kpelle society similar to the way that Binet wanted to use IQ tests.

Could we carry out such a program of research *in practice* ? There is no simple answer to this question, but it is useful to consider the obstacles. For some activities such as naming leaves or remembering riddles, it should be relatively easy to make the relevant observations because the Kpelle have already arranged for them : several researchers have described children's games that embody precisely these activities (Cole, Gay, Glick and Sharp, 1971 ; Lancy, 1977 ; Kulah, 1973). We could also test people's skills at constructing houses, weaving designs and forging sturdy hoes. However, from a Kpelle point of view, test of such skills would not be particularly interesting. The real stuff of using one's wits to get along in the world have been excluded.

This point was made very explicitly by a sophisticated Kpelle acquaintance of mine who was versed in the more esoteric aspects of Kpelle secret societies and medicine (or magic, according to American stereotypes). We had been talking about what it means to be intelligent in Kpelle society (the most appropriate term is translated as "clever.") "Can you be a clever farmer?" I asked. "No," came the reply. "You can be a hardworking farmer, or you can be a lucky farmer, but we couldn't say that someone is a clever farmer. Everyone knows how to farm. We use 'clever' when we talk about the way someone gets other people to help him. Some people always win arguments. Some people know how to deal with strangers. Some people know powerful medicine. These are the things we talk about as clever."

In this bit of dialogue we see an emphasis on activities that require social interaction as the arena where intelligence is an appropriate concept (among the Kpelle and many other non-technological groups, display of a good memory for use in discussions is often considered an important component of intelligence, e. g. Dube, 1977). This usage is quite consistent with Binet's analysis ; *it is those activities that differentiate among people* in terms of the way they manipulate information tha the Kpelle, like the French, mark intelligence.

However, once we reach this point, we face two important difficulties. First, the situations that we have selected for our study of Kpelle intelligence are exceedingly difficult to describe. Second, these contexts are very difficult to arrange. It is not enough to know riddles, everyone knows riddles. What is important about riddles is how they are used to get one's way with other people. Riddles are a resource to be used in a variety of social interactions where people's statuses and rights are at issue.

Consider the first difficulty. Bellman (1978) recounts an occasion when an elder member of a secret society told a long story about how he came to be a high ranking shaman. He followed this (presumably autobiographical) story with a long riddle, which was also in story form. A novice such as myself would have no way of figuring out what part of the story was true and I certainly would not have responded to the riddle as if its interpretation depended upon the autobiographical story; the two monologues appear to be about quite different topics. Bellman succeeds in demonstrating, however, that the riddle is closely linked to the autobiography. Not only are there formal, structural similarities (once one understands the basic categories of the relevant Kpelle belief systems). There is a rhetorical link as well. The autobiographical story actually represents a bit of self-aggrandizement by the person who told it. The man is claiming special knowledge and special power in a covert manner. The riddle reinforces the main point of the story (which raises the teller above his fellow shaman), giving the story "logical" as well as "historical" validity. The fact that listeners are constrained to agree with the riddle also gets them to agree, at least in part, with the message of the autobiographical story.

By almost any account, this man's autobiographical account plus riddle is a clever bit of behavior. It is exactly the kind of thing that our West African Binet ought to be sampling. But, at precisely this point, our cross-cultural thought experiment in IQ testing comes apart. As I have already pointed out, in order to construct a test Binet needed to be able to select a large number of items. But the "item" we have just described (very loosely) is not easily constructable. The participants in this scene were doing social work on each other; the shaman, in particular, was attempting to establish his preeminence using an account of his past history that would be difficult to check up on, a riddle whose structure was designed to reinforce his account, and his knowledge of his listener's state of knowledge concerning both the shaman's past and Kpelle social structure. This was one item; it was constructed by the subject, not the "tester." It is very difficult for me to imagine how to insure that a test includes one or more items "of this type." Furthermore, because the example's structure and content depend upon the special circumstances surrounding it, how could I insure that I would be able to present the test to the subject since it was the "subject" who did a lot of the presenting in the example I have described.

Here the contrast with Binet's situation is very strong. Like Binet, we have proceeded by figuring out what sorts of activities differentiate people according to some notion of what it means to behave intelligently. Unlike Binet, the activities we need to sample in West Africa to accomplish this goal lead us into domains that are *systematically absent from Binet's tests*. These domains involve interactions among people in which flexibly employed social knowledge is of paramount importance. They are not domains of hypothetical knowledge; rather, they always involve some real operations on the world, operations that require a great deal of care simply to describe. We have no good notion

of how to make such activities happen in a manner analogous to the way that teachers make vocabulary tests and multiplication problems happen. Furthermore, even if we solved all these problems, we would have no real theory of the psychological processes that our subject engaged in. Such problems have not been studied by cognitive psychologists.

On both practical and theoretical grounds, then, it appears virtually impossible to come up with a way of testing Kpelle intelligence in a manner really equivalent to what we understand to be intelligence tests in our society. So long as we restrict our attention to Kpelle culture, this conclusion should not cause much consternation. After all, the idea of a West African Binet is rather absurd; Kpelle people have managed to pass on their culture for many years without IQ tests to help them select clever children and give extra assistance to the dull.

SOME IMPLICATIONS FOR THE NOTION OF A CULTURE-FREE TEST

Our characterization of what one has to do to be clever in Kpelle culture and what it would take to sample such cleverness in a test must be discomfiting for anyone who imagines that they can construct a culture free test of intelligence. Imagine, for example, that by some quirk it was our imaginary Liberian Binet who constructed the first IQ test, and that other West African tribal people had refined it. Next imagine that American children were posed items from the West African test. Even items considered too simple for Kpelle 8 year-olds would cause our children severe problems. Learning the names of leaves, for example, has proven too difficult for more than one American Ph. D. (Bowen, 1964). Our children know some riddles, but little use is made of such knowledge in our society except for riddling, which would put them at a severe disadvantage on more "advanced" items.

If our children were forced to take a test constructed by a West African Binet, we might object that these Kpelle-derived items were unfairly biased toward Kpelle culture. If the eventual incomes of our children depended in any way on their ability to interpret Kpelle riddles, we would be outraged. Nor would we be too happy if their incomes depended upon their use of their own riddles as rhetorical devices. At the very minimum, we would want a *culture-free test* if real life outcomes depended upon test performance. However, what kind of test is a West African Binet likely to dream up that we would consider culture-free? It would not involve a set of drawings of geometrically precise figures, because Kpelle, a pre-literate group, do not engage in much graphic representation and they have no technology for drawing straight lines. It would not be recall of lists of nonsense syllables or even lists of words, because there are no corresponding activities in Kpelle adult life. We might try a memory test like recalling all of one's family, but here the Kpelle, who teach their children genealogies, would have a distinct advantage: what is the name of your grandmother's father on your father's side of the family? In fact, if we run down the list of presumably culture-free items that our experiment on Kpelle IQ testing turned up, we would almost certainly find none of the sub-tests that have been claimed as culture-free tests of intelligence in our society. The reason is very simple; our West African Binet, having scientifically sampled *his* culture, would have come up with items that reflect valued activities and that differentiate people in *his* culture, while Binet and all his successors, have come up with items that do the same job in our culture. *They are*

different kinds of activities.

The only way to obtain a culture-free test is to construct items that are equally a part of the experience of all cultures. Following the logic of Binet's undertaking, this would require us to sample the valued adult activities in all cultures (or at least two!) and identify activities equivalent in their structure and frequency of occurrence.

I probably do not have to belabor this point further. The simple fact is that we know of no tests that are culture-free, only tests for which we have no good theory of how culture affects performance. Lacking such a theory, we lack any guidelines that would permit us to specify clear connections between cultural experience and performance.

RETURN TO FIRST PRINCIPLES

Our imagined study of cross-cultural test construction makes it clear that tests of ability are inevitably cultural devices. This conclusion must seem dreary and disappointing to people who have been working to construct valid, culture-free tests. But from the perspective of history and logic, it simply confirms the fact, stated so clearly by Franz Boas half a century ago, that "mind, independent of experience, is inconceivable."

The historical experience of anthropologists has led them to consider it axiomatic that the abilities you choose to sample have to be drawn from an analysis of indigenous, culturally organized activities. Because different cultures emphasize different kinds of activities, the valued abilities will differ. From this point of view, a test that is equally valid across cultures would be a test that sampled some domain of activity that occurs in roughly the same form and same frequency in the cultures being compared. While it is possible, in principle, to identify such activities, they may not be of much use for the purposes of ability testing in the tradition of IQ tests. Many psychologists and anthropologists have asserted that some core set of experiences is common to all cultures by virtue of the fact. Such assertions are at the heart of such major systems as those constructed by Sigmund Freud, Jean Piaget, and Abraham Kardiner, to name just a few important figures who have studied this problem. But it is simultaneously asserted that everyone, irrespective of culture, comes to master those basic activities common to our species. Piaget, whose work is most closely associated with the development of intellectual skills, explicitly assumes that there will be universal acquisition of basic understandings of the physical and social worlds because of universal constraints on behavior common to all cultures. There are no existing data to refute this assumption.

However, both our 19th century anthropological forefathers and 20th century scholars such as Piaget readily admit cultural differences *associated with particular domains of activity*. Tylor, Spencer and other 19th century cultural evolutionists focused on differences traceable to technology. Piaget believes that special institutions and technologies of cultural transmission, such as the modern school, produce culturally determined cultural differences. So long as we restrict ourselves to specifiable domains, it is possible to rank cultures and, consequently, rank the intellectual achievements of individuals from different cultures within those domains. So, for example, we can rank cultures in the sophistication of their means of communication (from oral cultures, to literate cultures, to those possessing electronic media) ; we can rank cultures in terms of the complexity of dance movements that people are expected to master ; we can rank

cultures in terms of the degree of urbanization that characterizes the lives of their members, or the degree of rhetorical skill in institutionalized settings that they require.

Any time we engage in such domain specific comparisons, we can expect cultural differences in the abilities that individual culture users will have developed to achieve the required level of proficiency. Americans will be expected to deal more effectively with graphic symbols than Kpelle or Balinese. But if we chose dance movements as our subject matter, the opposite ordering of culturally linked proficiencies is certain to emerge. In either case, from an anthropological perspective, we would have no illusions that our tests of ability were culture fair. Why should we? After all, if we choose to compare people in domains where their experience differs, we expect mind to differ as well. That conclusion is certainly a basic legacy of 19th century anthropology.

Sticking to this point of view provides us with a powerful way of understanding the relation between IQ testing and social demands. We can recognize the school as an institutionalized setting designed to provide children with massive practice in activities that are useful and valued in our society. IQ tests sample school activities, and therefore, indirectly, valued social activities, *in our culture*. In so far as such tests are really used to insure that all children master the required skills, such tests would have to be considered extremely useful. However, in so far as such tests act as screening devices giving access to some people and not to others, without any commitment to insuring that all achieve the level of proficiency required for full participation in the adult life and access to the resources available to adults in our society, their initial purpose has been subverted and must be re-examined. This will be no easy task, since there is little current agreement on the intellectual skills needed for performing in most adult occupations.

With respect to the current American testing scene, the following observations seem most certain.

The discussion of cultural bias in American testing has been carried out on an extremely narrow base. Writers such as Arthur Jensen (1979) use that narrow base as a way of rendering plausible procedures for claiming that a test is "culture-fair" that are clearly inadmissible from the perspective of the logic of testing.

Jensen neatly demonstrates that tests are not necessarily statistically biased predictors of a criterion simply because they yield significantly different average scores for different cultural groups. Nor is it sufficient to point out that they contain items subjectively judged to be culture-specific (e. g., "How do you expect a kid from the ghetto to know what a violin cello is?" or "Ask that bright kid to tell you about the Man.") Subjective notions of culture specificity do not predict group differences in the way that cultural relativists would like ; those tests judged to be culture-specific often produce *smaller* group differences than tests assumed to be culture fair.

The problem with claims about the culture-free or culture-specific content of tests is that nowhere does Jensen or anyone else provide a demonstration that one test is more culture fair than another. The reason for this failure is simple : no one has set out to sample the experiences common to different cultures so that such a test could, in principle, be worked out. Instead, those who champion the Ravens Matrices or Kohs blocks as culture-free tests do so on the totally *ad hoc* assumption that the tasks are equally unfamiliar to all people concerned and the pattern of results they obtain ; it's a very

circular, as well as viscious, enterprise.

Our discussion in this paper applies equally to culture-specific and culture-general tasks : such tasks must be discovered by an analysis of the culture independent of test results, or claims for their culture fairness represent no more than the circular application of the unexamined intuition of the tester.

REFERENCES

- Bellman, B. L. *Village of curers and assassins : On the production of Kpelle cosmological categories*. The Hague : Mouton Press, 1975.
- Bellman, B. L. Ethnohermeneutics : On the interpretation of subjective meaning. In McCormack & Wurm (Eds.), *Language and Mind*. The Hague : Mouton & Co. 1978.
- Boas, F. *The mind of primitive man*. New York : Macmillan, 1911.
- Bowen, E. S. *Return to laughter*. New York : Doubleday, 1964.
- Cole, M., Gay, J., Glick, J. A., & Sharp, D. W. *The cultural context of learning and thinking*. New York : Basic Books, 1971.
- Dube, E. F. *A cross-cultural study of the relationship between "intelligence" level and story recall*. Doctoral dissertation, Cornell University, Ithaca, N. Y., 1977.
- Gay, J. *Red dust on green leaves*. Connecticut : Inter-Culture Associates, 1973.
- Gibbs, J. L. The Kpelle of Liberia, In J. L. Gibbs (Ed.), *Peoples of Africa*. New York : Holt, Rinehart & Winston, 1965.
- Kulah, A. A. *The organization of learning of proverbs among the Kpelle of Liberia*. Unpublished doctoral dissertation, University of California, Irvine, 1973.
- Lancy, D. F. Studies of memory in culture. *Annales of the New York Academy of Science*, 1977, **307**, 285-297.
- Spencer, H. *The principles of psychology, Vol. 5*. New York : D. Appelton, 1886.
- Tylor, E. B. *The origins of culture*. New York : Harper & Brothers, 1958.