Evolution of intelligence: Direct modeling of temporal effects of environment on a global absolute scale vs statistics

H.M. Hubey
Department of Computer Science, Montclair State University, Upper Montclair, USA
Evolution of intelligence

Direct modeling of temporal effects of environment on a global absolute scale vs statistics

H.M. Hubey
Department of Computer Science, Montclair State University,
Upper Montclair, USA

Keywords: Cybernetics, Intelligence, Computers, Brain

Abstract The social sciences are really the "hard sciences" and the physical sciences are the "easy" sciences. One of the great contributors to making the job of the social scientist very difficult is the lack of fundamental dimensions on the basis of which absolute (i.e. ratio) scales can be formulated and in which relationships could be realized as the [allegedly] coveted equations of physics. This deficiency leads directly to the uses of statistical methods of various types. However it is possible, as shown, to formulate equations and to use them to obtain ratio/absolute scales and relationships based on them. This paper uses differential/integral equations, fundamental ideas from the processing view of the brain-mind, multiple scale approximation via Taylor series, and basic reasoning some of which may be formulated as infinite-valued logic, and which is related to probability theory (the theoretical basis of statistics) to resolve some of the basic issues relating to learning theory, the role of nature and nurture in intelligence, the measurement of intelligence itself, and leads to the correct formulation of the potential-actual type behaviors (specifically intelligence) and dynamical-temporal model of intelligence development. Specifically, it is shown that the: (1) basic model for intelligence in terms of genetics and environment has to be multiplicative, which corresponds to a logical-AND, and is not additive; (2) related concept of "genetics" creating its own environment is simply another way of saying that the interaction of genetics and environment is multiplicative as in (1); (3) timing of environmental richness is critical and must be modeled dynamically, e.g. in the form of a differential equation; (4) path functions, not point functions, must be used to model such phenomena; (5) integral equation formulation shows that intelligence at any time t, is a sum over time of the past interaction of intelligence with environmental and genetic factors; (6) intelligence is about 100 per cent inherited on a global absolute (ratio) scale which is the natural (dimensionless) scale for measuring variables in social science; (7) nature of the approximation assumptions implicit in statistical methods leads to "heritability" calculations in the neighborhood of 0.5. and that short of having controlled randomized experiments such as in animal studies these are expected shrewdly due to the methods used; (8) concepts from AI, psychology, epistemology and physics coincide in many respects except for the terminology used, and these concepts can be modeled non-linearly.

1. Introduction
There is an old issue going back to Aristotle (who thought that slaves were slavish by birth), and which has become a heated debate in recent years by Burt, Spearman, Thurstone, Jensen, and Gould, having to do with the role of genetics and environment in intelligence, revived only a few years ago by Herrnstein and Murray (H&M). During this century the discussion has become
more "scientific" via the use of mathematical models and technique, the evidence consisting of tests, grading, and statistical analysis of such tests. Because of the importance of the history of the subject, the various incommensurable views adhered to by various parties, and sweeping breadth of discussion, the paper treads over known territory, some of it in standard/classical fashion and some with original twists, at the risk of boring some readers, in order to be accessible to the broad readership some of whom may not have any familiarity with some of the mathematical techniques. The attitude of the workers in this field, according to Herrnstein and Murray (1994) without oversimplifying, and without taking too long can be put into three groups:

Classic: Intelligence is a structure. Whether there is a single number, two or several is not as important as the fact that there's a structure to it, and this structure can be captured in a single number, which Spearman called g, general intelligence. Thurstone claimed about a half-dozen PMAs (Primary Mental Abilities). According to Vernon, they are hierarchical. According to Guilford there are 120 or so components in this structure.

Computational-AI model (revisionist): Intelligence is a process. This seems to be an evidently more modern attitude encompassing the information processing view. According to Sternberg there are three aspects of human information processing; the transducers or our sensory organs that change real world inputs into special forms for our brain, classifying the real world problems into groups, and actually making use of the apparatus in living (and hopefully being successful) in the real world via the use of the schemes of adapting (to the environment), shaping (the environment), and selecting (a new environment).

Scalar vs. Tensor (radical): There are different kinds of things called intelligence. For example, according to Gardner there are linguistic, musical, logical-mathematical, spatial, bodily, musical, interpersonal and intrapersonal forms of intelligence.

Of course, the phrase cognitive ability (CA) has now replaced intelligence and according to H&M, it is substantially heritable, apparently no less that 40 per cent and no more than 80 per cent. The importance of personal skills and emotional issues already clouds the definition of intelligence. Is it possible that all of these views have part of the truth and like the men who fought over what to do with their money without knowing that they all wanted to purchase grapes, they are fundamentally more in agreement as far as the facts are concerned just as men and women are more alike than unlike? Is there a unified view of which all of these are components?

1.1 Properties of intelligence: classical
One simple way of invalidating the results seems to be to deny the existence of race. The arguments are from biology; There's no scientific definition of race!
It's too silly to be of much use. For one thing, it won't stop the racists; another word will take its place. For another, biology is hardly in a position to be arguing about what is science and what is not since it is still rather low on the totem pole. And thirdly, the definition of race as given doesn't say anything more than what it's supposed to be: arguing that there's no such thing as beauty because it's only skin deep is silly. Who said it's supposed to be any deeper? We could, of course, see all kinds of beauty and in everything including in intelligence; indeed it exists everywhere. We might make up a simple table of words and phrases used in the literature for describing the intelligence or the CA debate as below (Table I):

> Man came first to the realm of the minerals, and from them he fell in among plants. For years he lived among the plants and remembered nothing of the vegetative state. In the same way he passed from realm to realm, until now he is intelligent, knowledgable, and strong. He remembers not his first intellects, and he will leave his present intellect behind. He will be delivered from this intellect full of avarice and cupidcity and see hundreds of thousands of marvelous intellects. Rumi (Chittick, 1983)

We can also use the standard terminology of nature-nurture, innate vs. cultural, but they all seem to boil down to a discussion of whether there is biological determinism. We might expand upon the standard arguments against the thesis that human intelligence is unfairly distributed to the different races by summarizing the arguments as making some version of the statement that, the IQ tests measure a

- **structure** (vector or tensor) AND a **scalar** should not or can not be derived from the tensor.
- **scalar** (single number) BUT cannot be used to linearly order or rank humans.
- component that's **race-based** AND is immutable.

The last part has been changed slightly from the ones expressed by Gould (1981). Some people might say 'genetically-based' (i.e. genetically inheritable) and slightly mutable. According to both detractors and proponents the IQ tests and their conclusions are about what is called biological determinism (BD); the idea that intelligence or cognitive ability/capacity is/are innate, intrinsic,

<table>
<thead>
<tr>
<th>Degree of freedom</th>
<th>Pole 1</th>
<th>Pole 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type-kind</td>
<td>Potential-virtual</td>
<td>Real-actual</td>
</tr>
<tr>
<td>Mechanism</td>
<td>Darwinist</td>
<td>Lamarckian</td>
</tr>
<tr>
<td>Sources</td>
<td>Genetic-biological</td>
<td>Learned-environmental</td>
</tr>
<tr>
<td>Elastic/plastic</td>
<td>Absolute-immutable</td>
<td>Relative-mutable</td>
</tr>
</tbody>
</table>

Table I.
inherited (biologically or genetically). A small problem revolves around the
definition of heritability. Heritable means capable of being inherited. Inheriting
has to do with coming to possess certain characteristics and not necessarily
genetically although it is often meant that way. So the fact that statistical
techniques such as correlation-regression, and analysis of variance have been
used to define inheritance or heritability means that somehow we are to assume
that we know exactly what it is and is clearly defined, but it is simply not the
case. It is as if we went to a doctor’s office complaining of our heart beating too
fast and he told us that we had tachycardia. He hasn’t diagnosed the problem
but only given it a name, and we should not be impressed. Several definitions of
heritability are possible (Table II).

Of these it is probably WAH and WAE are the most commonly held views
in opposition. So intelligence could be inherited, but only socially. Superficially
the basis of statistical inference and correlation-regression analysis seems
secure. Who would fight it? Most people would head for the intellectual hills
whenever faced with squiggly symbols of mathematics so the battle lines for
the Bell Curve would seem at a glance as if they resemble the Scopes monkey
trial with science once again about to triumph over its emotional opponents,
who naturally once again seem to be the ‘bleeding heart liberals’. It would be
strange to hear someone say that it is all for nothing but that is essentially what
it is about.

| race: A local geographic or global human population distinguished as a more or less distinct
group by genetically transmitted physical characteristics. |
| species: A fundamental taxonomic classification category, ranking after genus, and
consisting of organisms capable of interbreeding. |
| subspecies: a subdivision of a taxonomic species, usually based on geographic
distribution. |

| SAE (Strong Anti-Environmentalism) |
| all intelligence is due to heredity |
| and has nothing at all to do |
| with environment |
| |
| SAH (Strong Anti-Hereditarianism) |
| all intelligence can be accounted |
| for by environmental factors and |
| especially learning |

| WAE (Weak Anti-Environmentalism) |
| some (very little) has to do with |
| the environment and learning but |
| not much. Almost all intelligence |
| can be accounted for simply |
| by genetics |

| WAH (Weak Anti-Hereditarianism) |
| some of the differences in IQ |
| tests and thus intelligence is |
| due to hereditary factors but |
| this factor must be distributed |
| uniformly across all races so |
| that the differences between |
| the races is due only to |
| environmental factors |

Table II.
The theories of statistical testing (and CA testing debates) are replete with oblique axes, multicollinearity, orthogonal regression, covariant vs. contravariant tensors and to this we could add others such as rates of cultural vs. biological change. Some of this is done in detail in later sections. For a brief and intuitive tour of the relevant ideas we should turn to a short history of the vector vs. scalar theory of intelligence and thus to the pioneers of this century. The real vectors of mind, Thurstone reasoned, must represent independent primary abilities (PMAs). If they are truly independent they should be orthogonal (that is, perpendicular) to each other. But whatever these PMAs are, they are correlated; that is, they tend cluster. The problem is called multicollinearity in statistics. Not all sets of vectors have a definable simple structure. A random array without clusters cannot be fit by a set of factors.

The discovery of a simple structure implies that vectors are grouped into clusters and that clusters are relatively independent of each other; that is they represent, however inaccurately some aspect of some primary mental abilities, or PMAs. Thurstone identified seven of them: Verbal comprehension, Word fluency, Number (computational), Spatial visualization, M (associative memory), Perceptual speed, and Reasoning. Thurstone admitted strong potential influence for environment but emphasized inborn biology and also refused to reduce these to a single number, hence was an advocate of the structuralist school, it might be said. He claimed that Spearman's scalar g (general intelligence factor of some sort) was imply an artifact of the tests Spearman gave and nothing more. Spearman's retort was that Thurstone's PMAs were also artifacts of chosen tests, not invariant vectors of mind, which is also as true as Thurstone's claim.

1.2 Vector/tensor vs. scalar controversy: distance metrics & normalizations
Suppose we want to represent physical agility or physical capability of athletes from various different tests. Suppose we only use three tests; (i) endurance/stamina; (ii) reflex, reaction-time, and (iii) strength. How should we represent these three qualities (as quantities)? As the simplest such measure we can simply make three separate bits (i.e. zero or one) which will represent the possession or lack of the relevant property (such as a pass/fail grade) which we can write as 000,001,010,011,100,101,110, and 111 (see Figure 1). Or we can decide to give them grades in the normalized interval [0,1] for each of the three separate tests, and thus implicitly switch to using some kind of reasoning related to fuzzy logic or probability theory. Of course, we can easily increase the number of such tests to five or ten, and we can also increase the dimensionality of the problem but plotting more than 3 dimensions is very difficult. Hence, it is easy to deal with such high dimensional problems using only symbols and logic. To continue the example of 3 dimensions, we can make bar charts, pie charts or we can plot them on a 3-dimensional graph. Then we can represent each person as a point in three dimensions \( \{x, y, z\} \). We call such
ordered \textbf{n-tuples} or \textbf{vectors}. A vector is obviously a simpler case of a matrix. It is a 1 by n matrix. Matrices are also called tensors of rank 2, and vectors are tensors of rank 1. Therefore the ordinary single numbers are called tensors of rank 0 or simply scalars.

Consider the case of colors. Colors are produced from three so-called primary colors, Red, Green and Blue (RGB) or their complements, Cyan, Yellow, and Magenta (CYM) depending on whether an additive or subtractive process is used. No one would really argue that a color is not a single indivisible quantity if we think of it as something our perceptual/visual system is able to transduce. So then the natural question is whether a color is a single number, multiple numbers, a vector, a structure or a dynamic thing that causes our perceptual system to process the input data. It depends on our perceptual abilities and our knowledge. For sure it is all of them depending on what we want to do with it, and there’s no contradiction. As we know all the colors (for all practical purposes) can be obtained (additively) from the three basic primaries, Red, Green and Blue, (RGB) and Figure 2. The gray scale runs from black to white along the diagonal. The great advantage of using multiple dimensional space is the accuracy of such representations of much phenomena. We all know what colors are but they would be virtually impossible to explain to someone who was congenitally blind. If we did attempt to “explain” colors by explaining that “black is the absence of color and white is a mixture of all the colors” it is likely that the blind person would think of colors as what we call “gray scale”. We can write the primary colors as vectors

\[
\begin{align*}
\mathbf{r} &= \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} & \mathbf{g} &= \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \\
\mathbf{b} &= \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}
\end{align*}
\]  

(1)

Since a vector consists of ordered elements, the first entry refers to redness, second to greenness and the third to blueness. Thus the red vector \(\mathbf{r}\) has only a
1 in the redness-place and zeroes elsewhere. Similarly for the other primary colors, g, and b. We suspect, then, that the other colors will be some combination of these primary colors. What this boils down to is that we want to add different proportions of the primaries to create other colors so that we will multiply the primary colors by some number less than one (so that it is a small proportion) and then add them all to get some other color \( c_{\text{any}} \), so that

\[
\begin{align*}
\begin{pmatrix}
1 \\
0 \\
0
\end{pmatrix} + p_g \begin{pmatrix}
0 \\
1 \\
0
\end{pmatrix} + p_b \begin{pmatrix}
0 \\
0 \\
1
\end{pmatrix} = \begin{pmatrix}
c_{\text{any}}
\end{pmatrix}
\end{align*}
\]

(2)

where \( p_r = \) proportion of red, \( p_g = \) proportion of green and \( p_b = \) proportion of blue. If we had \( p_r = p_g = p_b = 0.5 \) we will obtain a gray since the diagonal of the color space that runs from black to white is called gray-scale. We can represent this particular gray as

\[
\begin{align*}
\begin{pmatrix}
0.5 \\
0.5 \\
0.5
\end{pmatrix} = 0.5 \begin{pmatrix}
1 \\
0 \\
0
\end{pmatrix} + 0.5 \begin{pmatrix}
0 \\
1 \\
0
\end{pmatrix} + 0.5 \begin{pmatrix}
0 \\
0 \\
1
\end{pmatrix} = c_{\text{gray}}
\end{align*}
\]

(3)

In the example above we saw the rules for scalar-vector multiplication and vector addition, but not vector multiplication. The final result for this particular gray is that it has 0.5 proportion of red, green and blue since those are the vector components. However, if we do make an analogy to the 3D space in which we live with the exception that the dimensions of color are not homogenous like our space dimensions, it is more likely to be understood.
better. For a more detailed look at color, see Hubey (1997). There is a simple
to obtain magnitude from the PA space (instead of using the Hamming
metric) by treating as the color space except that the meaning may not be
intuitive. Simply define

\[ PA = (eE_{2r} + \rho R_{2r} + \sigma S_{2r})^{\frac{1}{2}} \] (4)

For the special case of \( e = \rho = \sigma = e = r = s = n = 1 \) this is simply the
Euclidean distance metric that we use for our ordinary three-dimensional space
please see appendix A.1). Although it's just as obvious that the color vector is
being produced from the primary component colors our mind's eye sees a
single color. Indeed this is done all the time; the colors on the computer
monitors are produced directly by energizing the red, green and blue strips of
phosphors to varying degrees of intensity. The eye in the case of the high
resolution monitors (0.28 mm dot pitch) is unable to resolve the different
components and produces instead what we see as a single recognizable color on
the color wheel. The number does mean something. We can all see it. But
naturally we will not be able to assign a linear ranking since it's pointless. We
can see that the Euclidean norm of the color vector will be a section of a sphere
in the positive orthant but it could be one of an infinite number of colors on the
surface of this sphere. Making the analogy to colors what we can immediately see
is that our unaided intuition, if we only considered this color space to be a
homogeneous space like that in physics, would not be able to tell us that what we
perceive subjectively as color often does not seem to have any obvious connection
to the constituent components of the color vector, since we now know that what
looks to be a distinct "thing" is merely a shorter/larger wavelength in the visible
bandwidth of the electromagnetic spectrum. However, there is no doubt that any
given color can be comprised of the basis colors RGB. Therefore we have to
reason to insist that a vector created from the components of intelligence will
not possess intuitive properties totally different than the basis vectors. At the
same time, the scalar quantity obtained from the vector certainly is missing
much information. The real question is how different from each other are the
components of the intelligence vector.

However along the diagonal from black to white, we can indeed assign a
single scale the so called grey scale. And everyone will be able to visually
compare them. It will take some training to be able to estimate the color vector
components for various colors however in these days of computers it should
not be too difficult to be able to find a program with which to play around. And
indeed the results will be what we imagined above for the physical case, there
are differences and they are quite noticeable. So the whole idea of whether to
combine the components to produce a single number or to leave them alone may
not be much more than a matter of taste. In fact, if anything they should both be
done.
And different weightings should be used just to see what kinds of differences it would make. In the specific case of CA or PQ since the various alleged factors or components of the tests are or would be highly correlated and not independent as in the case of the three primary colors they would all be increasing more or less together and it would correspond almost exactly to the case of the grey scale, so there is something after all to what the classicists claim. Since they are correlated (i.e. tendency to increase or decrease together) then this resembles something like the grey scale and we can make use of this idea to comprehend what these tests purport to measure. So there's no serious difficulty with making sense of a scalar measure (i.e. a single number, say, Spearman's $g$). We can use analogical reasoning now to try to comprehend what this single number could mean, if we had a mind to produce such a single number. Indeed, it is an excellent example of the fact that although we can 'see' the grey number as something clearly related to black, we would not have been able to imagine that it is really being produced from red, green and blue. It is one of the miracles of the natural world; strange but true, just like finding order in randomness in chaos. But there is another simple way in which we can produce scalars from which we can get an idea of the colors. The problem of structure vs. process vs. multiple intelligences is a pseudo problem since the arguments are really about definition of intelligence. From the way it is explained it seems that by structure is meant really a state. In some ways the cognition view seems to be an attempt to solve the problem of intelligence by avoiding it, and the multiple intelligences view seems to be taking the vectors of mind view as is and refusing to go further. The only one that causes immediate grief is the difficulty of connecting the state view with process view since this is in general very difficult even in relatively well-tread fields such as thermodynamics.

2. Anomaly or contradiction? (data is evaluated according to a theory)
A more serious problem is the apparent paradox of the fact that we have the largest b/B for all animals (where b = brain mass, B = body mass) and yet brain size among humans doesn't seem to matter much from the evidence. Evidently either

(a) there is something analogous to flab for the brain so that massive brains don't necessarily imply high intelligence
or whatever intelligence is, the test doesn't measure it but rather a narrow set of skills taught to students who are expected to have this core knowledge just to survive and be a reasonably productive member of this society during this century.

or it's the connectivity that is important so that the more efficient connections may be present in the brains of some individuals who have small brains and are anecdotally said to have been 'smart'.

Figure 3.
(a) Global pattern but lack of correlation at local scale: A situation where using an absolute (ratio) scale would yield correlation over a large scale as expected but fail over small scales (magnitudes). The figure is not drawn to scale but is only meant to be suggestive. The simulated data points would be much more closely clustered (horizontally) in real life.
(b) The evolution of information content in genes and brains (after Britten & Davidson 1969; see also Sagan 1977). Compare this to (a).

Note: \( N \) in DNA nucleotide pairs per haploid cell.
It seems as if the correlation between the Encephalization Index (EI) (Eccles, 1989) and intelligence holds at large scales (i.e., global scale) and does not hold at small scales (local scale). We do definitely find that the larger the EI the more intelligent the species. Why then doesn’t the relationship hold at local scales? Superficially there could be two reasons; the tests (instruments) do not possess the resolving power required or that the relationship is not linear thus linear-correlation -regression (LCR) analysis does not divulge any information. However there are other reasons why the EI does not seem to correlate with intelligence at local scales (i.e., only for humans).

We do know that more complex organisms also have larger brain/body mass ratio (Britten and Davidson, 1969; Sagan, 1977)

2.1 Artificial (machine) intelligence perspective: form, mode, and type
As for the taxonomic structure of the skills that comprise what we call intelligence the first thing we note is that like a database there are different possible classifications, and that if they all seem to be just as attractive then they must be different conceptual views of the same thing which can possibly be all accounted for some day when we have better mathematical models. The standard models were reviewed in the beginning and we have yet more possibly taxonomies and also other pieces of evidence that points in the direction of a logarithmic scale. Some skills of problem solving are serial which would include what we call formal logic (and definitely its informal version that shows up constantly in verbal comprehension type questions), and some of the simple arithmetic (i.e., word) problems. Others the most obvious of which is spatial visualization require a parallel mode of processing. The visual [nonverbal and parallel] mode of thinking was probably best expressed by Einstein. Since the number of brain states (in analogy with computer science sense of the word i.e., say, a state of a set of flip-flops of a real machine or the internal states of an abstract machine such as a Turing machine) increases exponentially with the number of neurons, and we expect that ability, in some sense, also increases exponentially so that we should use a logarithmic scale. As for the complexity (in the sense of number of components of the number of operations a machine executes in algorithmic complexity) of the brain and expressive power of a language there are good reasons to think that they should be multiplicative and that there are trade-offs in time vs. space complexity, for languages (see Hubey (1994)). Going back to standard computer paradigms, if we concede that animals can think (although at some lower level) we must also concede that thinking doesn’t require language [if the few tens of words that animals can recognize is not counted as language]. There may be natural spaces in which to represent intelligence, which means that we may yet provide some kind of a structure to it. For example it would be possible to represent many of the ideas in terms of a simplified three-dimensional space whose axes are
Some kinds of questions require explicit knowledge such as mathematics, geography, verbal comprehension [grammar], and word fluency. Others are implicitly learned such as personal and interpersonal skills, physical coordination, and much of language. We might also call explicit knowledge much of what is taught in schools, and the implicit, what is learned without a formal education [which would include the so-called street smarts and also certain personality skills which would make for good manager or salesperson.] The last axis of the 3D space has to do with what might be called the difference in computation between batch vs. real-time or between I/O bound vs. compute-bound processes; it's really a combination of both. Into this last category (axis) would fall such things as bodily-kinetic intelligence of Gardner, musical talents (i.e. ability to play an instrument), being athletically minded, and perhaps some aspects of personality. Those involved in real-time programming know that it is a task of difficult constraints. Similarly, coordinating physical activity and mental tasks (i.e. as in team sports) is a rather difficult task i.e. high complexity. It is for this reason that music and dancing have calming effects; it stops the internal dialogue of Casteneda. It is for this reason that music might break some people's concentration but improve others'. We can try to include what should really be a fourth dimension, that is essentially a memory fetch vs. computation in this third dimension, but only to make its comprehension easier since representing more than three dimensions is very difficult except purely mathematically. The possible fourth dimension [only for the purposes of simplification and exposition], that of the difference between a compute-bound process vs. one of memory fetch in computer science would be the difference between a complex algorithm vs. that of table look-up. In the real world of humans, the table-look-up has the analog of word fluency, and perceptual speed [of Thurstone's PMAs]. Clearly it has to do with the organizational skills of the person, which naturally is about the organization of the knowledge of his brain, and hence his past which includes both formal and informal education. It is this which Thurstone probably calls M (associative memory). In Gardner's world view, this would get split into spatial (since perceptual speed might have to do with spatial resolution and manipulation of objects in space), and logical-mathematical would also be in this category. Since all memory in the brain seems to be associative, and analogically based, this particular component is probably what we might call efficiency in another setting instead of the intelligence debate and is probably what we are measuring along with some basic knowledge that we presume every human should know. Continuing with this idea we can see that things are often measured as a product of two variables in which one is intensive and the other extensive. For example, in
thermodynamics/physics work done is $\delta W = p \delta V$, entropy is $\delta Q = T \delta S$. The idea of intensive vs. extensive variables do have uses in many different areas. Training or education is probably something like $\delta t = x \delta T$ where $x$ is the intensity or quality of the training program and $\delta T$ the extensive variable which is the amount of time spent in it. Problem solving ability is $\delta \pi = e \delta K$ where $e$ has something to do with the inferencing mechanism or engine used, and $K$ the knowledge base, for despite all claims to the contrary and protestations, we cannot separate the two completely, at least in the human brain, and at least for the time being. Knowledge of the world comes from our senses, and our inferencing about the world at large comes from our observations. In fact, we can see the same ideas being used in scoring in gymnastics and diving. The score is calculated by multiplying the raw score (how well performed) by an inherent degree of difficulty of the routine of the dive. Hence the measurement is really about a product of an intensive parameter (organizational effectiveness of the brain or its efficiency) multiplied by an extensive parameter which is knowledge. Please see appendix A.3 on Path Integrals and their connection to these ideas.

2.2 Potential and its realization
In articles on intelligence (indeed almost any other characteristically human trait such as language) we often run into words which talk about human potential which has not been realized. It is often thought to be a single dimension in which the actual realization is simply a proportion of the potential (capacity). What seem like two poles of a continuum often turn out to be separate dimensions. The case of language turns out to be one of these. There are really two variables; capacity & existence of instruction. There is a window of opportunity for picking up language.

We see that we are dealing with a product of variables since it is only a product which can create this. And in this case the simplest approximation is just a logical AND. In other words, there must be both language-capacity (i.e. innate, inherited, potential) AND also there must be proper environment (i.e. instruction), so that language can be learned. The next level of approximation is simply using fuzzy logic concepts. As long as the potential is there (for example in mentally retarded children) and there is instruction, there will be some form of language. Indeed, IQ tests do measure language competence to various degrees and use it as a part of the test of intelligence. Combining the two, (i.e Knowledge Form, Computation Mode, Bound Type) and the concept of potential from physics we might try a potentials of form

$$\Psi = \alpha e^{F_{it} + p M_{it} + T_{it}}$$

$$Q = \alpha F_{it} M_{it} T_{it}$$
from which we can compute the vectors of the mind, and also derive single or multiple scalars using any of the ideas shown in earlier sections. The potential in Equation (6) is already multiplicative and Equation (5) becomes additive as \( \ln(\Psi) = \phi F^m + \mu M^m + \tau T^m \) after taking logarithms so that if we are interested only in adding up scores on various sections of the test without any compelling reason not to do so, the logarithm will relate these numbers to the potential. In the later case, the logarithm produces the standard form for linear regression. Without some data on what these mean it be pointless to speculate on the choice of functions however we should note that \( Q \) is multiplicative so that if any one of the components is zero \( Q \) will be zero, so that it already has built-in correlatedness for the components. It would tend to produce high scores for more well-rounded informal low level education [i.e. cognitive intelligence] whereas if there is a limit to what is possible because our brain is finite after all, then the high achievers would certainly be deficient in some areas and stronger in others which would be exaggerated by the exponential form of \( F \) given the right coefficients, so both forms are flexible enough for creative use. Even a simple multiplicative model is much closer to the truth than the standard linear regression models. Many things having to do with psychophysics is best modeled by a power law, and the sigmoidal functions, as pioneered by Rasch (1980) seem to have much success. Other sigmoidal models can be seen in Hubey (1987) and stochastic models in Hubey (1991a) and also below. We still have the problem of obtaining the actual/real from the potential (which is the concern of learning theory).

2.3 Functional view of the psychology/biology of learning, and intelligence

It's reasonably clear from all the evidence that whatever it is that intelligence tests measure whether it should be called the Intelligence Quotient or Cognitive Ability or Problem Solving Capability, or Problem Solving and Creativity Scale or whatever can be changed/affected via training, emotion, poverty in other words environmental influences. Even after scaling things correctly, we are still left with variation among humans. It might be argued despite the evidence that it still means something and needs an "explanation", in other words, some simple theoretical model. It is not difficult to produce a very simple model that hopefully will not do terrible injustice to the idea of intelligence. We know that our memory is associative. Memory events seem to be linked to other events and we can recall almost everything with some prompting. We might make an oblique reference here to artificial intelligence programs in which an inference engine is working on data so that we may liken problem solving to having an inference engine (naturally not necessarily localized in some part of the brain but possibly scattered about) fetching data and doing some kind of a search (breadth first, depth first or some combination thereof or something completely unknown to us yet). Of course it will take time to do all this.
Let us call the time it takes to do this \( T_w \) for complete search time, without implying that the search does not include conventional computation i.e. problem solving. Suppose now that over a period of time we’ve built up (via formal or informal education) a large bag of cheap tricks which is also kept in storage someplace. We can think of reasoning as analogical reasoning in which we solve problems via analogy to problems resembling the given problem in one or more dimensions and that we kind of keep a mental template of such solved problems in memory (which we might imagine is functionally being kept someplace separate from the rest of the memory). Thus if we first are able to find the template for the given problem in this memory of pre-solved problems [premem] we can ‘solve’ the problem much faster than we could have if we never had encountered problems of this type. The truth of the matter is that there are really no completely original or novel problems that can be presented at any of these tests. And solving some of them really revolves around guessing what the tester wants to get as an answer. Therefore the time to solve a problem if we can find an analogical match in our pre-solved memory is much shorter than if we treated the problem as completely original and tried to be creative in its solution in which case might never be able to solve it at all. The time to solve a problem, then having this highly-simplified two-tier memory system then drops to

\[
T_b = HT_p + (1 - H)(T_p + T_c)
\]

(7)

where \( H \) is a probability of finding this solution in premem [pre-solved memory]. This is essentially what is being referred to as “chunking” in learning theory and in artificial intelligence. Thus if we do find it there, the answer is very quickly found, which takes time \( T_p \). The time to find the solution if it’s not found in this pre-memory fetch is that time spent doing this plus the time spent actually solving it via supposedly original methods. Naturally, this simplification is so gross that we should not expect anything beyond the simplest kind of description of matching reality. First, there really is no such

![Figure 4.](image)

Memory levels: A simplified functional view of memory needed for solving problems and the role of learning.
thing as these two memories locatable anywhere in memory but there's no need for it; the connections must behave something like it. Secondly, we'd have a tough time solving very original problems; if anything the problem we have is in finding a good match for the problem at hand and trying to force fit couple of problems together or cobbling solutions from several such virtual templates; it is this efficient time that we've called $T_p$. In any case, the tests don't give us time to find the solutions but rather give us a fixed amount of time in which to solve such problems, so the assumption behind this idea is that we'll be able to solve less problems in this fixed amount of time if we cannot find many of them in our premem. In truth all of us who are alive have some small virtual memory in which already solved problems from life are stored (naturally not necessarily in some localized region of the brain) so that the time it takes for us to solve the problem compared to some hypothetical baseline of the finely-tuned problemsolving brain would be of the form

$$\lambda = T = T_s/T_p = H + (1 - H)(1 + T_c/T_p)$$  \hspace{1cm} (8)

We should note that this solution time, $T$, is equal to $H$ if $H = 1$ and is equal to $1 + T_c/T_p$ if $H = 0$. The case of $H = 0$ corresponds hypothetically to a situation in which we are faced with a problem for which we have no handles. In this case the factor $T_c/T_p$ is something that corresponds to the inherent originality of the problem, at least to the subject. We should suspect that $\lambda$ should be a large number since very few people [almost no one] are actually creative but rather partially creative; we may cobble together solutions to new problems by combining several old ones. We solve large problems by cobbling together solutions to a bunch of smaller component problems. The process is iterative, and hierarchical since the same types of solutions can be used at different scales, hierarchies or levels. Most people probably cannot even do that unless they've been trained to do so. In any case, highly educated people, especially those who've studied mathematical sciences will have high $H$ values since they probably have already solved symbolic problems of the type found on tests many times over. Similarly, questions such as “what is the opposite of...” will be easier for those children raised by parents who are highly literate than those living in “tarzan neighborhoods”. We should really consider not $T$ but another quantity $\tau = 1/T$ if we want to consider the values as normalized. We then have

$$\tau = \frac{1}{H + (1 - H)(1 + \frac{T_c}{T_p})} = \frac{1}{H + (1 - H)(1 + \lambda)}$$  \hspace{1cm} (9)

The plot shows that for $H = 1$, $\tau = 1$ since at that time we have reached maximum efficiency since every problem presented to us in the test is already present in our pre-solved memory and we need only to fetch the answer. We
should also note that the most rapid increases occur for large $I$, which is exactly as it should be since it implies that the ratio of the learned solutions to the searching/groping for solutions using all ingenuity and creativity is high, meaning that if the problems are of the type that would be very difficult to solve without being exposed to problems of this type, then the steepest increases come near $I = 1$ when we can find the solutions in the pre-solved memory. It's possible that the large brain individuals may be capable of more original lines of thought, capable of more creative lines of thought, have more memories built in. It's also possible that the so-called intelligent beings such as mathematicians or novelists, or philosophers were merely one-dimensional experts in small domains and managed to score high on these tests particularly because they were trained for these tests. In particular the tests might overemphasize classification which is a large component of education, especially in the 'soft sciences'. It is said that an expert knows everything about nothing and the generalist knows nothing about everything. This is simply an example of trade-off as can be observed in many fields (Hubey, 1996), and also Appendix A.2. Most people would naturally have to fall somewhere in between. In yet another sense we can consider the effect of $I$ plotted against the amount $S_p/S_c$ where $S_p$ and $S_c$ are proportions of the memory devoted to the two different types of problem solving modes and their associated memories, where we've assumed that there must be some kind of a parameter $\Omega$ which has to do with the organization of the brain. If knowledge is organized so that there's a method to the solution searching mechanism instead of being a cut & try method that an unsophisticated person might attempt, the probability of finding the answer (or something close to it) in the faster premem $I$ will increase. Hence we might think of $\Omega$ as a kind of efficiency of the brain as far as
its organization goes. It's also possible that this could point to overorganization in the sense that it will be good only for solving the types of problems given on such tests. As can be seen if there was absolutely no efficiency raising mechanisms or learning by experience, hence no localization of memory (i.e. associativity) then the increase in H should be about linear with S_p/S_c. There should be a higher rate of increase of H with S_p/S_c if the learning mechanism was efficient instead of simply being rote training. In all likelihood memory (that is, neural net) organizes itself in some manner which is captured in an extremely simple way by these equations. The early methods of solving problems are much closer to the parts of the triune brain (MacLean, 1973; Jerison, 1973) so that they become automatic means or fall-back methods, and thus the increase in the likelihood of finding the solution to problems such as those given in various IQ/CA tests greatly increases performance. This "organizational efficiency" of the brain has been captured in the single parameter Ω. Other thoughts on functional descriptions of memories of living entities include the procedural vs. declarative memory (Squire, 1983), working and reference memory (Olton, 1983), and associative and recognition memory (Gaffan and Weiskrantz, 1980), which like the present work is borrowed directly from computer science. For trade-off type relationships in many fields of science, and epistemology, see Hubey (1996).

3. Mathematical analysis of proposals
The previous section was a purely functional view of the role of learning in problem-solving, but IQ/CA is not supposed to be learned but innate/hereditary/genetic. If intelligence cannot be learned, what exactly, then, is IQ? To answer this we must first ask what intelligence is. IQ is a normalized version of intelligence. The question has obviously been asked and answered in different ways in the past. In binary form the answer is the Turing Test. To know what intelligence is in nonbinary form we should try to

![Figure 6. Effect of localization of memory and specialization: With early learning there is more efficient organization of the brain for certain types of tasks and thus leading to higher Ω than for late learning](image)
delineate its properties. Some of this was already done in the beginning in the literature review. In this section we can try to produce answers from other points of view, ignoring the previous section and re-starting new thread by examining the standard arguments but evaluating them from different perspectives. Historically, the brain/mind was always described by using the highest technology available as metaphors for understanding and what we are attempting to understand or describe is a function of the brain/mind. The mind was likened to clockworks, then the telephone switch, then the digital computer and finally artificial neural networks. The memory part was likened to holograms, and the associative memories of computer science are still used as analogies. The computational paradigm is still rampant, and the concepts of state and process come from this view. However since the brain/mind is a very complex thing, there is yet one more analogy we can make, and that is to databases which have different conceptual (often called logical) views. The multiple view perspective is taking hold these days even in operating systems. Since analogies are always single dimensional, it is not surprising that something as complex as the human brain/mind (the three-pound universe) can be seen to be like so many things. Since we don’t yet understand it whole but only its parts, we can liken ourselves to the story of the four blind men and the elephant. There are other questions we can ask regarding its properties. Is it an extensive property or an intensive one? Is it like temperature or pressure (i.e. an intensive function) or is it like volume/capacity/mass/internal energy?

The answer to both is that it is probably a product of both! Not only is the problem solving ability a function of some kind of an efficiency of neurons or organization of the brain but also of the pure mass or amount of neurons. If it were not so, animals such as reptiles would be as intelligent as humans. On the other hand if we claim that since we are only considering humans, and since the brain masses all fall into the same range, we should consider this constant, then we still have to deal with whether IQ is intensive or extensive purely from the consideration of whether it depends on knowledge (extensive) and also on some kind of an efficiency of processing or creativity in solution finding (intensive). Therefore we still cannot escape the bind of choosing one or the other. It is most likely a function of both and hence it must still be multiplicative function, aside from the problem of being a path function and not a point function. On the basis of the foregoing we can find at least four serious problems with the attempts by which psychologists so far have tried to capture the idea of intelligence, aside from the ones that have already been discussed in the literature and earlier in this text.

- What kind of a quantity is intelligence? Is it binary or measurable on some scale? What kind of a scale is appropriate? Is it an ordinal, interval, or an absolute (ratio) scale?
- Is it an additive function of its constituents, the most important ones for purposes of simplification being hereditary(nature) and environment
(nurture)? Or is it a multiplicative function? Is it logarithmic function, an exponential function or a polynomial function of its variables?

- Is it a vector/tensor function or a scalar?
- Is it a point function, or a path function? In other words is it a state or a process? Is it a quality or a quantity? Is it an extensive variable or an intensive variable?

We all recognize that genetic influence can be spread diffusely among many genes, and that genes set limits to ranges; they do not provide blueprints for exact replicas. In one sense, the debate between sociobiologists and their critics is an argument about breadth of ranges. For sociobiologists, the ranges are narrow enough to program a specific behavior as the predictable result of possessing certain genes. Critics argue that ranges permitted by these genetic factors are wide enough to include all behaviors that sociobiologists atomize into distinct traits coded by separate genes. Gould (1981), 329.

It's clear that all of these questions are not independent of each other but related to one another. If this thing called intelligence is to make any sense it should be comprehended and comprehensible in a broader context. It is paradoxically true that sometimes one can find solutions to problems by generalizing them and looking for more general solutions since that enables us not only to locate the phenomena in its proper space relative to related ideas or objects but also allows us to use more data as evidence to better grasp the constraints to be imposed on the phenomena. This intelligence scale should encompass and allow us to measure intelligence of fleas, as well as that of chimps, humans and also machines.

Common sense says that the scale should be logarithmic in order to accommodate the vast differences in intelligence but also because many laws in psychophysics are power laws. Logarithmic transduction of inputs allows for a greater range of sense perception without a proportional increase in the size of the organs. Furthermore, if this scale is to be something like the temperature scale, then absolute zero should belong to something like viruses or simple computer programs. Furthermore ideally this scale should be an absolute/ratio scale instead of simply an interval or an ordinal scale. A highly mathematical treatment of the subject of scaling going back to Campbell (1920) can be found in Suppes & Zinnes (Luce et al., 1963).

3.1 Heritability: why is the "intelligence function" not additive?

First problem with the Linear Correlation-Regression Models (LCRM) is that it is highly unlikely that intelligence is an additive function of environment and heredity since additive means logical OR and not AND. So therefore the verbal expression that intelligence is a function of both environment and heredity is being twisted out of shape as soon as we try a linear additive model. As is well known, AND is represented as multiplication, and not necessarily only in
bivalent logic or even fuzzy logic but even in modeling via differential equations, for example in the nonlinear Lotka-Volterra models the interaction is multiplicative (see Appendix A.5). Various types of infinite-valued AND functions can be found in Hubey (1998). The sigmoidal function is produced quite naturally in the nonlinear differential equation modeling of forced binary discrimination of phonemes in Hubey (1994).

Biological determinism is fundamentally a theory about limits... Why should human behavioral ranges be so broad, when anatomical ranges are generally narrower?... I conclude that wide behavioral ranges should arise as consequences of the evolution and structural organization of the brain. Human uniqueness lies in the flexibility of what our brain can do. What is intelligence, if not the ability to face problems in an unprogrammed (or as we often say, creative) manner? Gould (1981):331.

Additivity implies that the environmental and hereditary components are grossly substitutable for one another which is simply untrue. No amount of teaching will make a chimp into a human. There is no question that the model should be multiplicative. The model cannot be additive since additivity logical translates to OR and nobody would really dispute that environment and heredity are not grossly substitutable for one another. If it were so, we could be teaching calculus to dogs by enriching their environment to make up for their genetic deficiency. The coefficients no longer mean what they meant in linear regression. If we are looking for the magnitude of variation of intelligence with the factors the two cases give fundamentally different results because if we have \( I = f(E, G) \) then

\[
\delta I = \frac{\partial f}{\partial E} \delta E + \frac{\partial f}{\partial G} \delta G
\]  \hspace{1cm} (10)

For the linear case

\( I = \Phi + \varepsilon E + \kappa G \)  \hspace{1cm} (11)

the differential (i.e. variation)

\( \delta I = \varepsilon \delta E + \kappa \delta G \)  \hspace{1cm} (12)

For the nonlinear case

\( I = \alpha E^e G^h \)  \hspace{1cm} (13)

the variation/differential is

\[
\delta I = (\alpha E^{e-1} G^h) \delta E + (\alpha h E^e G^{h-1}) \delta G
\]  \hspace{1cm} (14)

As can plainly be seen from the form of the multiplicative (i.e. AND) dependence the powers of \( G \) and \( E \) essentially determine the sensitivity of the
intelligence to variations in environment and heredity. For the linear case the respective coefficients do determine the sensitivity of intelligence to the factors, but for the nonlinear case (which is the correct case) the respective coefficients no longer mean what they meant for the linear case. The model must be multiplicative. (See Appendix A.7 for some paradoxes.) The simplest such model accounting for environment and heredity would be of the multiplicative type which is interpreted as a logical-AND (i.e., conjunction) Therefore the linear regression could be done via using the logarithms would have the form

$$\ln(I) = \ln(\alpha) + e\ln(E) + h\ln(G)$$

(15)

Immediately, we would see that all the numbers that were measured would get smaller and hence the variances. However, that is not the only problem (see Appendix A6 for correct computation of variation and Appendix A4 for the conditions on the functional form). The argument that the present testing methods and models are only for “human level intelligence” where the linearity is valid does not hold water for there are standard mathematical methods to deal with such approximations. We simply expand the function in a Taylor series and attempt to regress about some point which we may claim is some average human level genetic and environmental condition and that the function is approximately linear about that point. For example, if we suspected some general form such as $I = f(E, G)$, then we can expand

$$I = f(E_h, G_h) + (E - E_h) \frac{\partial f}{\partial E} E=E_h + (G - G_h) \frac{\partial f}{\partial G} G=G_h + \ldots$$

(16)

which for Equation 13 above is

$$I = aE_h^\delta G_h^\gamma + (E - E_h)(aeE_h^{e-1}G_h^\gamma) + (G - G_h)(ahE_h^{h-1}G_h^\gamma) + \ldots$$

(17)

Rearranging terms and simplifying we obtain

$$I = \Phi + \Delta E + \Lambda G$$

(18)

where $\Phi = aE_h^\delta G_h^\gamma (1 + e + h, \Delta = aeE_h^{e-1}G_h^\gamma, \Lambda = ahE_h^{h-1}G_h^\gamma$. In order to make Equation (16) linear we dropped the higher order terms in the Taylor series to obtain Equation (18). However, the linear correlation-regression analysis computes the value of the constants $\Phi, \Delta$ and $\Lambda$ which parameters are no longer indicative of the effect of the self-variables since they are now functions of the other variable. In order to offset this dependence we would have to use the normalization $E_h \cdot G_h = 1$ thereby computing the coefficients $\alpha(1 + e + h), \alpha e$ and $\alpha h$ in the linear regression. We can then solve for $\alpha, e$ and $h$ from the three equations. If we do solve for these coefficients in terms of the regression values $\Phi, \Delta$ and $\Lambda$ we obtain the results:
\[ \alpha = \Phi - \Delta - \Lambda \]
\[ e = \frac{\Delta}{\Phi - \Delta - \Lambda} \]
\[ h = \frac{\Delta}{\Phi - \Delta - \Lambda} \]

If we had, say \( \Phi = 2 \) and if \( \Delta + \Lambda = 1 \), then the above works out only to rescale the parameters since we would then have \( e = \Delta \) and \( h = \Delta \) so nothing would really change. If \( \Phi < 1 \) we'd obtain negative correlation and we cannot allow \( \Phi = 1 \) since the numerator would then be zero. However, if we had used another scale, say the one in use right now (i.e. \( E_h = G_h = 100 \)) everything would not work as above. Something which depends on a particular choice of interval scaling for its truth cannot be correct. We do not know if the present IQ scaling is meant to be an interval scale or an absolute scale. It is through problems like this that Kelvin's research led to the postulation of an absolute temperature scale. (Please see appendix A1 and Appendix A6. For more on fuzzy logic and differential and the meaning of multiplication and nonlinearity, please see appendix A. 4.)

3.2 Problem of dynamics in measurement and attribution of causality

We should note there is another complication since the real complexity of the problem is in the dependence of the variables on one another since they can be functions of one another. For example, if we are traveling in an airplane from Maine to Florida starting up at around 9:00 AM and taking measurements of the ambient temperature, the rate of change of temperature we'd measure is that not only of the spatial variations in temperature (north-south) and also the temporal variations since the air would start to warm up after the sun comes up and will be reaching a peak say around noon. Since we have the temperature \( q = q(x(t), t) \) where \( x \) is the distance traveled starting form Maine then the rate of change of the temperature measured (recorded by the instrument across time) is

\[ \frac{dq}{dt} = \frac{\partial q}{\partial t} + v \cdot \frac{\partial q}{\partial x} \tag{20} \]

where \( v \) is the velocity of the airplane. The first term is the purely temporal rate of change of the temperature (which is due to the warming of the earth because of the sun), and the second in which the term \( v \cdot \frac{\partial q}{\partial x} \) appears [which is the spatial change], multiplied by the velocity of the airplane gives the change due to both the actual spatial variation and the rate of sampling of this spatial thermocline. For the case of measuring intelligence (whatever it may be) we don't know that the variables we have selected are really independent. For example suppose we have \( y = y(M(t), V(t), t) \) [where \( M = \text{mathematical}, \ V = \text{Verbal}, \text{and } t = \)
training i.e. formal or informal education. We know that verbal ability is important because without it we can’t even give these tests. But are we sure that mathematical/symbolic/quantitative reasoning is not important for verbal comprehension? What exactly is the relationship between the two? In terms of the neural networks underlying these they are both handled by neurons, although there is lots of evidence of localization of speech, spatial reasoning etc. (for example Gazaniga, 1985; Sperry, 1988; LeDoux, 1977). However our main concern now is in mathematical formulation of the problem. Since speech and visual ability are developing in infants simultaneously, in all likelihood three-dimensional spatial comprehension and its verbal articulation probably go hand in hand although people seem to start early into developing some modes more than others, for example, spatial orientation, verbal fluency, physical development.

In the study of any scientific discipline it is necessary, in the beginning stages, to use words whose precise meanings may not be defined, but are accepted as defined in an intuitive sense as starting points. On the basis of the ideas and concepts derived from these basic terms, a theory begins to develop and then it is possible to retrogress and give precise quantitative definitions to the words and terms defined only verbally. Perhaps, the best example of this process is in the field of thermodynamics. Concepts such as heat, temperature and pressure were properties only physically felt and intuitively understood. After thermodynamics was put on a theoretical footing, the concepts of temperature, heat and pressure were defined operationally (mathematically on the basis of the developed micro (kinetic-statistical) theory of thermodynamics. Hubey (1979).

4. Putting it all together: effect of learning and timing of learning on potential
Many things which are accepted to be a part of the “natural” (how this word is abused probably will take a book to explain) growth/maturatation of humans are all due to learning. For example, in very early ages, we are told that it is quite “natural” for children to have pretense play and to invent objects and people. In all likelihood, this is due simply to the fact that the infant still has not made the strong differentiation between sleep/dreams and wakefulness. The child falls asleep in one place and wakes up in another (for example in the car, at the beach, or in someone’s arms). This is probably no more mysterious at this time than being in one wonderland (in sleep i.e. dreaming) and then waking up to another reality in another place. At the same time, if it is talking to dolls or toys or dogs, it is still learning that somethings are alive and move on their own accord, some are toys and are run by batteries and that somethings that move (i.e. toys or animals) do not speak. Another stage in growth/development is when it still does not understand for example the concept of picture so that if we tell it to “do a truck” it might mimic driving it instead of drawing the picture (Gardner, 1991). But of course, does the child at that age understand that small
iconic representations of objects which it sees on TV, or a book got there by various means such as a camera or being "drawn" by other human beings? It is simply ignorance, nothing more. If someone draws a picture in front of his eyes (not a bad picture since it might not be able to make the connection well at that stage) it might think that the pictures on TV are also drawn or it might think that there is a little guy inside a Polaroid camera like the German peasants during the last century who thought there were horses inside the locomotive. For the case of measuring intelligence (whatever it may be) we don't know that the variables we have selected are really independent. For example suppose we have $\psi = \psi(M(t), V(t), t)$ [where $V$ = Verbal, $M$ = mathematical, and $t$ = training i.e. formal or informal education], then the variation in the potential is

$$\delta \psi = \frac{\partial \psi}{\partial M} \delta M + \frac{\partial \psi}{\partial V} \delta V + \frac{\partial \psi}{\partial t} \delta t = \psi_M \delta M + \psi_V \delta V + \psi_t \delta t \quad (21)$$

where we denote the partial derivatives by subscripts, so that if we wanted to know the change in this potential with respect to training (which would naturally affect the measured intelligence) we'd need to compute the total derivative

$$\frac{d\psi}{dt} = \frac{\partial \psi}{\partial M} \frac{dM}{dt} + \frac{\partial \psi}{\partial V} \frac{dV}{dt} + \frac{\partial \psi}{\partial t} \quad (22)$$

It's possible that $d\psi/dt = 0$ since we cannot think now of how $\psi$ could change if neither $M$ nor $V$ changes (assuming that these are the only factors/variables we've identified). In the general case, naturally, there'd be more variables. But in truth things are more complicated; it may be more like $\psi = \psi(M(V(t)), V(t), t)$. It's obvious that we can't even get across the problem let alone the solution without language so that $V$ will definitely affect $M$. In this case we have

$$\frac{d\psi}{dt} = \frac{\partial \psi}{\partial M} \frac{dM}{dV} \frac{dV}{dt} + \frac{\partial \psi}{\partial V} \frac{dV}{dt} + \frac{\partial \psi}{\partial t} = \psi_M \frac{dM}{dV} \frac{dV}{dt} + \psi_V \frac{dV}{dt} + \psi_t \quad (23)$$

where $t$ is a proxy for the environmental richness. It seems at this point that we can get stuck in infinite regress since if $M = M(V)$ we may then need to write $V = V(M) = V(M(V)) = V(M(V(V(M))))$ if we cannot separate the influence of $V$ directly and via $M$ and if $V$ is also a function of $M$. In some problems we can measure this, say, in industry analysis in economics. In any case, we can see the effect that this will have on computing the gradient of the potential to derive the vector function. Or we might have a more complex case such as that of having a potential of the form $\psi(M(V(t), V(t), t)$ or even $\psi(M(V(t), V(M(t), t))$. There is even a more serious objection to linear correlation-regression analysis. We can see immediately that new memories are built on
top of old ones, and that learning to solve problems is just as good as being creative, and in many cases, the learning eventually outstrips creativity and that is naturally the reason why the testing for IQ stops at adulthood since everyone has already pretty much learned what there is to learn, and that if the same trends continued, we should be asking questions on algebra, trigonometry and calculus on the IQ tests given to adults, say college students or graduates. The fact that it is not done is a testimony to the simple fact that the tests also test for knowledge.

Furthermore the earliest memories should count more heavily, even in using the standard IQ tests since new memories are built on old ones, and thrusting someone into a new socioeconomic (SE) class is not the same as having someone in that SE class since childhood. In fact, there is probably a lag of several generations at least, and probably centuries as can be seen in the long cycles of histories of countries and empires. Therefore we already know that IQ is a \textit{path function} and not a \textit{point function}. Again it is not a \textit{state function} but a \textit{process function}, where state and process are used in a more general sense than in computer science or in psychology, or in mental testing (please see Appendix A.3). The earliest such models come from thermodynamics. And it is also from thermodynamics that we have the ideas of extensive vs. intensive properties of systems. The standard example of a path function is the length of a curve on a plane. The standard example in physics comes from thermodynamics in which the heat rejected or the work done of a system depends on the path that the process took and not a function of the end states of the process. We can thus surmise that intelligence will be a function of the time, and the path that the environment variable, say, $E(t)$ took during this time.

\begin{quote}
\textbf{Advantages of mathematical or analytical models-}\textbf{unambiguity, possibility of strict deduction, verifiability by observed data-are well known and make them highly desirable in systems engineering. This does not mean that models formulated in ordinary language (e.g. verbal descriptions) are to be despised. A verbal model is better than no model at all, or a model which because it can be formulated mathematically falsifies reality. Indeed theories of enormous influence such as Darwin’s Theory of Selection, were originally verbal. Much of psychological, sociology and even economics today is still descriptive. Models in ordinary language, therefore, have their place in system theory. The system idea retains its value even where it cannot be formulated mathematically, or remains a “guide” rather than being a mathematical construct.}
\textbf{Hubey (1979)}
\end{quote}

Thus in addition to the problem that what we purport to measure, say, verbal skill $V$ which may be a function of mathematical or spatial skills $M$ or vice versa, we now have a bigger problem of the form of the function itself. The significance of this cross-dependence of variables is obvious if we think of the fact that some proposals have been put forward that the richer \textit{environment itself is a function of genetics}, i.e. $I = f(E(G), G, t)$ (see for example Plovin and Bergeman (1991) for a review). Obviously, this is true on a global scale and the
derivatives of this function will have to be calculated in the same way as that of Equation (23). More on this can be found in Appendix A6 and in the conclusion section. We can approach this problem a little differently: if we have say some learning ability, \( L \), we can see that it will in all likelihood be a function of time, since the earliest years are the most important and by old age very rarely can people retain their mind's elasticity of their early years. However, here intelligence is a function of this learning ability for which we use time, \( t \) as a proxy. Furthermore an enriched environment is essential, and the earlier this rich environment is provided the better it is, so that we can attempt to surmise the form of the functional dependence of intelligence on environment. We have
\[ I = f(G, E(t), L(t)) = f(G, E(t), t) \]
where we have accepted that \( t \) is a proxy for the learning ability, and that \( E(t) \) and \( G \) are representative of or proxies for environmental and genetic variables, respectively.

We expect that we should have \( f_G > 0 \) since with a more enriched environment we expect increases in intelligence. Similarly we expect that \( f_t > 0 \) if people are measured on the same scale since problem solving ability should increase with age (please see Appendix A.4). Note that we are not discussing IQ which can be obtained from intelligence as a result of normalization. The intelligences of people have been increasing over the past half a century or so. One of the reasons, of course, is that the environment itself has been changing. Not only has the educational level gone up, but the environment itself (i.e. the standard of living) has gone up, and thus children are being exposed to more things and are getting better care both healthwise and nutritionwise. Consequently, not only do IQ tests measure knowledge [albeit claimed not to measure knowledge at all but some kind of "innate/genetic" capacity] but whether it means an intensive variable or an extensive variable or a product is not clear in the literature since it hasn't been discussed with respect to any model except in terms of regression or correlation coefficients. The tests also measure a quantity which is a function of the path (the history of the individual in a particular changing environment). In other words, the words process and state are not necessarily to be understood only in the sense made popular by the emergence of the digital computer as the metaphor of choice among philosophers and scientists working in the intelligence/knowledge field but

![Evolution of intelligence](image)

Figure 7.
Realization of potential:
The potential (heredity) for speech is there for all humans but if the window opportunity for learning language passes language cannot be learned. Meanwhile, although instruction (environment) is given to animals they cannot learn to speak. The dividing lines are arbitrary and merely suggestive.
rather in a more general sense in which thermodynamics made popular. We should construct a path function for the dependence of intelligence on environment. The simplest path function is the length of a curve, and it is an integral. Taking a cue from this we may try a very simple function of this form for intelligence to be of form:

$$I(E, t) = \int \eta(E(t), t) dt$$  \hspace{1cm} (24)$$

However it would be preferable to derive such a function from more basic considerations instead of producing it out of thin air, like the standard linear correlation-regression analysis. Otherwise we could be accused of behaving like the man who was looking for his lost keys in his yard because there was more light there. Even worse, we could be accused of behaving like the little boy, who, given a toy hammer, discovered that everything looks like a nail. Since the process of acquiring intelligence (as measured in some fashion by standard intelligence tests) is a dynamic process we should turn to differential equations. In the case of a simple differential equation, a first order ordinary differential equation, given by

$$y'(t) + b(t)y(t) = f(t)$$  \hspace{1cm} (25)$$

has the solution

$$y(t) = e^{-\int_0^t b(\tau)d\tau} \int_0^t f(\tau)e^{-\int_0^\tau b(s)ds} d\tau + y(0)e^{-\int_0^t b(\tau)d\tau}$$  \hspace{1cm} (26)$$

We note that for a constant a, it reduces to

$$y(t - \tau) = \int f(\tau)e^{-a(\tau-\tau)}d\tau + y(0)e^{-a(t-\tau)}$$  \hspace{1cm} (27)$$

so that the Green's function (the exponential function which is the kernel of the integral) of the convolution integral is, in a sense, a weighting function, since it assigns an exponentially decreasing weight to the earlier forces that affect the system. In contrast, the weighting function that should be used for the effects of environment on intelligence should give greater weight to the earlier times since it is now common knowledge that brain damage can occur even in the womb due to effects of simple things like smoking cigarettes. Since the fastest growth for the brain, as well as the body occurs during the early years, it is a foregone conclusion that it should be so. In this case, in addition to the fact that the basic model should be multiplicative, the environmental factors should be a path function, something that accounts for the effects of the environment at different phases of development.
Clearly the differential equation model which had "feedback" can be modeled as a "black box" and we can find out from the inputs and outputs how the internal mechanism of the "black box" works. The black box models in the social sciences are strawman arguments, not against behaviorism but against science. This area of "identification" of the system (i.e. black box) and prediction based on its is a well-developed science. The "black box" model (i.e. Equation 27) is "time-invariant" in that the parameters of the differential equation are not time-dependent. In the real world, the behavior of intelligent entities changes with time; it is a part of learning. These changes require nonlinear differential equations and this topic is discussed briefly in appendix A.6 (Meaning of Nonlinearity). We know that as children grow, not only do their bodies develop but so do their brains and minds. A simple example of growth is given by the differential equation

\[
\frac{dy}{dt} = k(1 - y(t))
\]  

(28)

We assume that the growth rate of the child is proportional to the size yet to be achieved; that is, it grows faster when it's smaller because the size yet to be achieved is much larger than in adolescence when it has reached close to its adult height. The brain also grows at similar rates and we can take this equation to be a simple model for the development of intelligence for a start. As is well known the solution consists of an exponential approach to a final constant value. The coefficient g can be seen in this plot to control the rate at which the child approaches its final adult height, and would grow faster for larger values of k. Of course, this is a simplified model and does not take into account the fact that there are spurts in growth rates around puberty. If anything a large g would indicate a precocious child, especially if its intelligence were to increase at the same rate. At this point we need to consider other global effects in what to expect. On a large scale, as can be evidenced every day, we see that except for humans all other animals seem to have a limit of intelligence and capability/capacity beyond which they cannot advance. We already know that a multiplicative formulation is needed, therefore we need to combine this idea with the dynamics of intelligence. On large scale over time we expect to see what is shown in Figure 9(a).

It would seem on a global (large) scale that intelligence is definitely genetic. No dog will ever talk and no chimp will ever do calculus. So then why do the statistical tests give results that intelligence is 40 per cent to 80 per cent genetic instead of more like 99.9999 per cent? As a simple example of the kind of mathematical equation for the above we can try

\[
I = A_x(1 - e^{-kt})
\]  

(29)

The real question then becomes, exactly what kind of function of heredity or environment are the parameters \(A_x\) and \(k\)? As can be seen from the plot and the
Figure 8.
Various "Black Box" Models: It is thought by some that the last two models with a feedback loop from the environment is not a part of the "black box" method of science and cannot be handled by standard mathematical tools. Of course, that is not how it is practiced in real life. See Appendix 17

(a) Introspective Psychology: adopted from Ledoux [1996]

(b) Standard Black Box Model and Behaviorism [Ledoux, 1996]

(c) Cognitive Science: Functionalism and Computationalism [Ledoux, 1996]

(d) Feedback Model & Loops [Hubey, 1994, [Naylor & Sell, 1992]

(f) Mathematical "Black Box" Formulation for Linear Dynamical Systems
equation, the parameter $k$ determines how fast the intelligence of the subject increases in time toward its potential which is apparently mostly genetically determined. But if we examine these plots at small scales (or higher resolutions) say only for humans then we see something like this (again simplified). Of course, in reality both $A_x$ and $k$ fluctuate and although the plots do not show them a typical sample function could crisscross across others. In any case, now that we look at it at small scale (and high resolution) we have other means of interpreting the differential equation that gave rise to this

\textbf{Evolution of intelligence}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure9}
\caption{Increase in intelligence of various species after birth on an absolute intelligence scale. The slow increases up to some limit are typical of exponential curves. (see appendices 3 and 4). The initial condition is not really zero but is drawn this way for simplicity. (9b) Variations in the parameters of the intelligence model for a given species. It is evident that both parameters $A_x$ and $k$ are functions of both heredity and environment. The arrows show the “track jumping” behavior of Head-Start type programs in which change in the environment puts the child into a different sample function (i.e. path).}
\end{figure}
The rate of increase of intelligence is proportional to the intelligence yet to be achieved with $\gamma$ being the constant of proportionality. The intelligence limit on the whole is determined genetically but it acts as a kind of an attractor of expectation from the child. In other words the difference $(A_x - x)$ can easily be thought of as motivation. Does $k$ then denote the genetically determined factor (i.e. a rate of increase)? Since we already have much evidence that this constant (is not constant but varies) can be changed with more attention and greater quality and quality of teaching and practice it cannot be a purely genetic factor either. It could be genetically determined on the whole but it is also a factor of both heredity and environment. So then we are led toward the complete model which could be of form:

$$I = aE^{\alpha E G^\gamma} (1 - \lambda e^{-\lambda E^\gamma G^\gamma})$$

This is clearly the solution of the differential equation for intelligence

$$\frac{d}{dt} I(t) + \lambda E^\gamma G^\gamma I(t) = \lambda aE^{\alpha E G^\gamma + \gamma}$$

which although simple and linear still has basically all the right ingredients to be a model of dynamical learning in the nature-nurture environment. From the solutions of first order ODEs of this type (e.g. Equation 25 through Equation 31) the coefficient of $I(t)$ (i.e. $\lambda E^\gamma G^\gamma$) determines the rate of increase therefore it is the part that represents the interaction of the environment with genetics. The limit intelligence ($\lambda E^{\alpha E G^\gamma}$) is achieved eventually but if this were completely independent of the coefficients $\alpha$ and $\lambda$ it would mean that all this interaction has nothing to do with intelligence and two other coefficients representing something else (i.e. $\gamma$ and $\lambda$) determine final components of intelligence. The important part is that the multiplicative interaction of $G$ and $E$ is modeled. Variations on this theme can be seen below and in Appendix A.4.

4.1 Much ado about nothing?

In mathematical modeling, it is really the equations that talk. However the meanings of these equations has been discussed throughout the exposition. By changing the scales to the natural (absolute) scales, and by making the global intelligence/behavioral parameter multiplicative with genetics factors and by examining the behavior of this function in the neighborhood of human level behavior we can unify much of the work done on such topics as has been done. We can improve the model above by making appropriate changes, pointing out its relationship to the constraints that must be satisfied by such models, and by connecting it to the standard analysis of such problems in the literature. We see
clearly that Equation (31) is simple linear (but dynamic) realization of a more
general form

\[ I = \beta G^h \Phi^p(E(t), t) \]  
(33)

where \( \Phi(E(t), t) \) is a path function examples of which are given in Appendix
A.A. The reasoning to obtain the differential Equation (32) was already given,
but there are criteria/constraints that it must satisfy to be a good representation
of the genetics-environment interaction. Since \( E \) is not constant but heredity is
fixed at conception (at least at present) a more general (and slightly different)
version of Equation (32) is

\[ \frac{d}{dt} I(t) + \lambda G \gamma E^e(t) I(t) = \lambda \alpha G^h + \eta E^{e+e}(t) \]  
(34)

which by virtue of Equation (26) has the solution

\[ I(t) = \lambda \alpha G^h + \eta E^{E + e} \int_0^t E^e(\tau) d\tau \int_0^t E^{E + e}(s) e^{\lambda G^h} \int_0^t E^e(\tau) d\tau \]  
(35)

There is yet more to the power that this simple linear differential equation
hides. Integrating it once and rearranging terms we obtain the integral
equation

\[ I(t) = K(t) - \lambda G \eta \int_0^t E^e(\sigma) I(\sigma) d\sigma \]  
(36)

where

\[ K(t) = \alpha \lambda G^h + \eta \int_0^t E^{e+e}(s) ds. \]

The interpretation of this equation is exactly what is claimed by most
researchers in the field, namely, that intelligence at time \( t \), that is \( I(t) \) is a
function of the past interaction of intelligence with environment summed up
over time from time zero to the present time \( t \). The \( K(t) \) term is also a
multiplicative function of environment and genetic interaction and its
position is reminiscent of the differential equation formulation in that it
seems to be some ultimate potential for a given environment and genetic
makeup (for all humans) toward which all humans grow. Obviously,
Equation (32) can easily be cast in integral form as above and has the same
interpretation. This Equation (36) is a simple version of a more general integral
equation, for example, in Appendix A.4. A more convenient form of this
equation (especially for purposes of testing, which is discussed in Appendix
A.6) is in Appendix A.4.

The solution of this integral Equation (36) [and the equivalent differential
Equation (32)] is given in Equation (31)] and in it we can see that after a
sufficient amount of time elapsed, the effects of environment and genetics wears off so that the limit

$$I = \alpha E^* G^h$$

(37)

is reached [where we have made the change of variables $ee \rightarrow e$ and $h\eta \rightarrow h$ from Equation (31)] which is the original multiplicative form that was posited based on fundamental reasons valid for all intelligent organisms. Indeed, the complexity of the real world is much beyond what can be captured by these linear and deterministic equations. More on this train of thought can be seen in Appendices I3 and I4. Even if we did stick to these linear models (i.e. Equations (34) or (36)) we still have to consider that an ensemble of such/ would be needed, one for each person with its own parameters. That would mean that we'd have to consider the parameters E and G as random variables thus turning Equation (34) into a stochastic process in which we would compute the probability density $\rho(l, t)$ of the process. For simple cases we can obtain solutions for example in Appendix A.8. That such methods are the wave of the future for highly complex problems of the social sciences has been argued in detail in Hubey (1996) and examples of simple solutions can be seen in Hubey (1993) and complex ones in Helbing (1995). Since the intelligence has been normalized to unity, then Equation (37) is really another expression of the relationship of environment to genetics, which can be written as $E^* G^h = \text{cons}$. In other words, the nonlinear formulation as in (37) is not necessarily the only one, but rather an example. There are other formulations possible, for example

$$E^* = C e^{-\mu G} \quad \text{or} \quad E e^{G} = \text{cons}$$

(38a)

$$e^{E^* G^h} = \text{cons}$$

(38b)

$$e^{E^* + G^h} = e^{E^* G^0} = \text{cons}$$

(38c)

Whatever the case, a linear approximation via the Taylor series in the neighborhood of human level genetic endowment and cultural/environmental achievement will all lead to the same linear approximation results. Since the linearity is obtained from an approximation the numbers are only good in the neighborhood where $E + G = 2$ (see appendix A6). But then we have

$$E \approx 2 - G \quad \text{or} \quad \delta E \approx \delta G$$

(39)

which leads inexorably to the conclusion

$$\delta E^2 \approx \delta G^2$$

(40)

Since these have to add up to the total variation of unity, it is not a surprise that the environmental variance or genetic variance hovers around 0.5 in studies (Rowe, 1994; Rushton, 1997; Plomin and Daniels, 1987; Plomin and Bergeman,
1991). Similar to the situation in economics (see Appendix A4) this fact is really an indication that the socioeconomic and technological systems and the educational systems that support such systems are created via complex interactions so that we (humanity) work near our optimal limits. It is interesting that similar results hold in the case of the production functions of economics theory, for example, the Cobb-Douglas type, which is multiplicative as here, that the “share of capital” and “share of labor” in the production process and function are about 0.5. In the case of the production function, the numbers are merely a reflection of how the socioeconomic system is setup. For example, in LDCs where they must buy machines from other more developed countries, the “share of labor” is much less than the “capital’s share”. The reason that in advanced societies, the shares of labor and capital are about equivalent merely reflects the fact that the machines which are in the production process are built by other workers in industry and machine costs reflect the salaries and wages of those involved in the production of these machines. Furthermore it is a sign of the power of workers since if capitalists were really that powerful, they could conceivably pay pittances to all the workers, and claim large shares of the profits for themselves. In a similar vein, the reason the heritability is about 50 per cent is really a reflection of the way tests are created and is an indication of the importances that society attaches to skills in various ways. It would be quite easy to create tests in which word recall, reading would be rated low, and mathematical-symbolic-logical capability rated greatly and skew the results of tests so that the heritability results are more skewed than they are now.

If these intelligence tests given today were given to our ancestors 5,000 years ago they probably would have scored about the same as those in less developed countries for most of what passes as intelligence is really knowledge of the shared environment which in advanced societies is shared through the popular mass media organs such as television, and propagated through our educational institutions. If anything, the scores less than 0.5 are a testimony to the unequal environment in our societies. Similar views have been voiced by others, for example very strongly by Lewontin (1975).

To show explicitly the effect of the nonlinearity on one of the infamous variables of cognitive ability studies, we can show that in the linear (additive) case such as

\[ I = \alpha + eE + \gamma G \]  

(41a)

the heritability coefficient calculations are rather easy since

\[ \delta I = e \cdot \delta E + \gamma \delta G \]  

(41b)

and therefore

\[ H = \frac{\gamma^2 \delta G^2}{\delta I^2} = \frac{\gamma^2 \delta G^2}{e^2 \delta E^2 + \gamma^2 \delta G^2} = \frac{\gamma^2}{e^2 + \gamma^2} \]  

(42)
The last step was obtained assuming that the virtual variations (displacements) are equal i.e. \( \delta E^2 = \delta G^2 \) and by assuming the crossproduct term to be zero as is usually done when the interaction term variance is ignored. We can do this because we want to know how much of the variation in \( I \) is due to unit variations in \( G \) and \( E \). Behind this approximation are really stochastic differentials. In other words, alternatively, we may treat the variations to be random quantities and then average over the ensemble in which case assuming that the variations \( \delta E \) and \( \delta G \) are independent (which is the assumption used when ignoring the \( V_{GE} \) in the standard calculations in ANOVA) the crossvariance is zero. The result is clearly what is expected, and that the heritability is really the ration of the variation due to genes to total variation. What is hidden here but assumed is that both \( G \) and \( E \) are measured on the same ratio scale since if it were not so the equality of the small variations would not be assumed. For the nonlinear dependence of intelligence on both the environment and heredity as given by

\[ I = \alpha E^\gamma G^\gamma \]  

the variation/differential is

\[ \delta I = (\alpha E^\gamma G^{-1}) \delta E + (\alpha \gamma E^\gamma G^{-1}) \delta G \]

To compute the heritability as usually done we would have to divide the variation due to the variation due to genetics by the total variation. In this case we compute it (ignoring the cross-product as is usually done to be

\[ H^2 = \frac{(\alpha E^\gamma G^{-1})^2 \delta G^2}{(\alpha E^\gamma G^{-1})^2 \delta E^2 + (\alpha \gamma E^\gamma G^{-1})^2 \delta G^2} \]

\[ H^2 = \frac{\delta G^2 (\alpha^2 E^2 G^2 G^{-2})}{\delta G^2 (\alpha^2 E^2 G^2 G^{-2}) (\alpha^2 E^2 G^2 G^{-2}) + \gamma^2 G^{-2}} = \frac{\gamma^2 G^{-2}}{\delta E^2 G^2 G^{-2} + \gamma^2 G^{-2}} \]

The result can be put in final form which can be analyzed for semantics.

\[ H^2 = \frac{\gamma^2}{\epsilon^2 \left[ \frac{G^2}{E} \right] + \gamma^2} \]

Clearly, it is now even more obvious that both \( G \) and \( E \) must be measured on a ratio scale (i.e. absolute scale) since now the ratio shows up explicitly in the calculations. Simple analysis of variance calculations based on a linear (additive) model of the influence of genetics and environment in which they can be used as substitutes for one another, is clearly false. The fact that only equations for variance are almost always used in the discussions such as
\[ V_P = V_E + V_G + V_{GE} \]  
\[ V_G = V_A + V_D + V_I \]

(47a)  
Evolution of intelligence  

(47b)

hides the fact that these ad hoc derivations rely and are based on linear/additive models as shown above. If \( E \) is measured on a scale, even if it is a ratio scale, in which it is, say, 10 times larger than the scale on which \( G \) is measured, the product will make the contribution from the environment seem small so that the heritability coefficient will get larger. To see the devastating effects of nonlinearity on computations of heritability as is usually done, we can examine a simple case of nonlinearity in which \( I = aEG \). Substituting \( e = \gamma = 1 \) into Equations (45) or (46) we obtain

\[ H^2 = \frac{G^{-2}}{E^{-2} + G^{-2}} = \frac{E^2}{G^2 + E^2} \]  

(48)

It is impossible under these conditions to claim that \( H^2 \) really measures (the genetic component) heritability! Obviously, to obtain true heritability we must compute \( 1 - H^2 \). Furthermore it can be shown that if the dependence on the environment and genetics is \( I = e^{eg^2} \) the results for \( h^2 \) (or \( h^2 \) as appropriate) are still the same as above. In the case of the dynamical equation or the solution [Equations (31) and (32)] the equation for \( h^2 \) is still the same form as above, therefore the heritability calculations based on the linear model are incorrect. Furthermore, such calculations need to be made on measurements based on a ratio scale so that arbitrary scales for socioeconomic factors cannot be used. We can obtain similar results via Taylor series. Expanding Equation (43) about \( E_0 \) and \( H_0 \), we obtain

\[ I = E_0H_0 + (E - E_0) \frac{\delta I}{\delta E}_{E_0,H_0} + (H - H_0) \frac{\delta I}{\delta H}_{E_0,H_0} + \cdots \]  

(49)

\[ \frac{(I - I_0)^2}{I_0} = \frac{(E - E_0)^2}{E_0H_0} \frac{H_0}{E_0} + \frac{E_0}{H_0} (H - H_0)^2 \]  

(50)

where \( I_0 = E_0H_0 \) which is the average/normal human intelligence. Obviously then we can identify \( \delta H^2 = (H - H_0)^2 \) and \( \delta E^2 = (E - E_0)^2 \). It is clearer from these derivations that the previous analysis was basically the equivalent of analysis of variance, and thus the results have been demonstrated for a variety of ways.

If these intelligence tests were given to our ancestors 5,000 years ago they probably would have scored about the same as the semiliterate peoples in the less developed countries for much of what passes as intelligence is really knowledge of the shared environment which in advanced societies is propagated through the popular mass media organs such as television, and through our educational institutions. If anything, the wide variations on these
tests are a testimony to the unequal environment in our societies. Similar views have been voiced by others, for example, Lewontin (1975). In terms of the processes which give rise to such scores, there is basic agreement among many workers in the field, except for some lingering confusion. For a more detailed exposition of the ideas, see Hubey (1996). The fundamental concepts are shown in Table III below. Most of the time by “qualitative” people mean “not well understood” because many intensive variables are quite easily quantifiable. It is an unfortunate accident of history and sloganeering that a word like “quality” has come to be the basis of a word like “qualitative” which is used in opposition to quantitative to disparage the physical sciences. The most correct version of all of these is the intensive-extensive dichotomy (Hubey, 1996) which is what the psychological division of associative vs cognitive/conceptual signifies, as can be easily seen via extrapolation from the AI concepts of knowledge-base and the inference-engine which operates on it. In humans both of these are stored in the brain using neurons. The earlier some of the inferencing mechanisms (i.e. intensive variable) are learned the more they become a natural part of the human reasoning process similar to talking and walking and the more easily they are able to masquerade as intrinsic/genetic factors. The increase in the probability of finding similar problems already solved in memory (section 2.3) greatly increases performance (Figure 5). The earlier these are learned, the more efficient the brain organization for problem-solving (Figure 6). Therefore more accurate measurement these effects requires models in which time is explicit. This also explains why brain size does not correlate more strongly with intelligence tests. Problem-solving techniques, whether learned informally during early childhood or formally in school present themselves in studies as “intelligence”. It is for this reason that more difficult questions are not asked on such tests and especially to adults. Many people in the physical sciences and mathematics would score very high on such tests, but then the learned component would be very obvious to every researcher and layperson alike. However, when fundamental concepts learned early in childhood and which adds to the efficiency of brain organization are asked on such tests we are instead left with “controversy”. It is for this reason that some researchers have put forward ideas such as musical talent, body intelligence and the like. This argument misses the point if one can retort “are

<table>
<thead>
<tr>
<th>Philosophy</th>
<th>Physics</th>
<th>Artificial intelligence</th>
<th>Psychology</th>
</tr>
</thead>
<tbody>
<tr>
<td>quantitative, e.g. extensive knowledge-base, e.g. any associative, e.g. memorization of field-specific database of data and information</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>almost any variable, e.g. volume, entropy cognitive/conceptual, e.g. fluid</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>physical science Intensive inference-engine, e.g. intelligence, chunking by</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>qualitative, e.g. variable, e.g. breadth-first vs depth- first searching strategies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>most “soft pressure, temperature</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sciences”, philosophy knowledge</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table III.
there music neurons” or if music is noncomputational. Clearly, music is also computational (Johnson, 1997). In the past some researchers took refuge in such arguments are creativity, and originality (in anti-AI arguments) and musical-kinetic intelligence (in anti-math orientation of test questions). However, the validity of the argument stands if it is about the lack of natural dimensions and the weighting of the distance metric in n-dimensional space. (Appendix A.1).

At this point in time unless tests which are stringently controlled can be given, and tests which explicitly take into account nonlinear interaction of genetics and environment, there is not sufficient reason to attribute differences in performance on standardized tests to genetic differences which is not to say that what the questions test for are not important to society. If however, motivation is a key factor, one might ask why people with PhDs cannot learn 2 semesters worth of calculus or physics over a period of 50–60 years.

In general the differences between humans and other animals (say chimps) in all measurable behavioral characteristics are likely differences of degree and not differences of kind unless there are definite physiological constraints. This means that the interval [0,1] is the natural absolute scale of measurement, and the maximum will be achieved by our species which provides for normalization at the upper end of the scale. Furthermore the natural kind of relationship is multiplicative, which can still be tested using standard methods using logarithms in which there is a trade-off between order of magnitude and the nonlinearity of the logarithmic transformation. If Taylor series approximations are used to obtain linear relationships to be tested, the necessity of using the natural scales is obvious, otherwise the interaction of the different factors cannot be separated from each other.

References


MacLean, P.D. (1973), A Triune Concept of the Brain and Behavior, University of Toronto Press, Toronto.


Plomin, R. and Daniels, D. (1987), "Why are children in the same family so different from one another", Behavioral and Brain Sciences, 10, pp. 1-60.


Further reading


Appendix A.1: Scales or levels of measurement

Before we try to measure or normalize quantities we should know what kinds of measurements we have. They determine if we can multiply those numbers, add them, rank them etc. Accordingly measurements are classified as: (i) Ratio scale, (ii) Interval scale, (iii) Ordinal scale, or (iv) Nominal scale.

According to Thurstone, the decision to use factor analysis as a primary method implies a deep ignorance of principles and causes (1981: 316)

Absolute (ratio) scale: The highest level of measurement scale is that of ratio scale. A ratio scale requires an absolute or nonarbitrary zero, and on such a scale we can multiply (and divide) numbers knowing that the result is meaningful. The standard length measurement using a ruler is an absolute or ratio scale.

Distance: Probably the most common measurement that people are familiar with is that of distance. It is such a general and common-sensical idea that mathematicians have abstracted from it whatever properties it has that makes it so useful and have extended it to mathematical spaces so that this idea, in fact, used and useful in the previous ideas of measurements. The requirement that the concept of distance satisfies is this:

\[ d(x, z) = d(x, y) + d(y, z) \]  \hspace{1cm} (A.1)

The concept of “distance” or “distance metric” or “metric spaces” is motivated by the simple concept illustrated Figure A1.

If we substitute from the figure above we can see that the distance from LA to NYC can never be greater than the distance from LA to some intermediate city plus the distance from that intermediate city to NYC. The most commonly known distance measure is the Euclidean metric which most of us have seen in some form or have heard about as the Pythagorean Theorem. In the plane the Euclidean metric follows directly from the Pythagorean theorem as shown.

Hamming distance: Hamming distance is the number of bits by which two bitstrings differ. For example the distance between the bitstring 1111 and 0000 is 4 since the corresponding bits of the two bitstrings differ in 4 places. The distance between 1010 and 1111 is 2, and the distance between 1010 and 0000 is also two.

Phonological distance: distinctive features. In phonology, the basic primitive objects are phonemes. They are descriptions of the basic building blocks of speech and are usually described in binary as the presence or absences of specific characteristics such as voicing, rounding, frication, plosivity etc. Since we can represent these as bitstrings the Hamming distance can be used to measure the distance between phonemes. One simple way to describe a subset of phonemes (vocalic phonemes) especially of the Turkic, Mongolian and Uralic languages is via the Hubei Ordinal Vowel system (Hubei, 1994) below Figure A3.
Such hypercubes can be constructed also for sets of consonants. We need higher dimensional or fractional spaces to represent them all on the same graph or hypercube or same space. In order to represent them all there are several ways.

**What's a bird?** The concept of distinctive features can also be used in conjunction with fuzzy logic in artificial intelligence to describe (or define) objects, such as a bird, fruit or a chair. For example, a set of simple properties such as "has feathers", "is bipedal" and "flies" is generally sufficient to define a bird for intelligent entities (such as humans). We can use a simple Hamming distance or try a Euclidean distance.

**Interval scale:** However, not everything that can be measured or represented with integers (or real numbers) is not a ratio/absolute scale.

**Ordinal scale:** The next level on the measurement scale is the ordinal scale, a scale in which things can simply be ranked according to some number but the differences are not valid. In the ordinal scale we can make judgements such as A > B. Therefore if A > B and B > C, then we can conclude that A > C. In the ordinal scale there is no information about the magnitude of the differences between elements. We cannot use operations such as +, −, × or ÷ on the ordinal scale.

![Distance Metric Diagram](image)

\[ d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \]

**Figure A2.**

**Distance Metric:** One would not expect, if the above were the real space of our existence, that the x and y dimensions would be measured on different scales, i.e. x in feet and y in meters before using the formula above for distance. It would be an unthinkable error, but we do not know how we should combine scores on many tests in psychology because the dimensions such as Space/Length, Time, Temperature and Charge of physics do not exist.

**Figure A3.**

The ordinal vowel cube (Hubey, 1994)
**Libert scale:** It is possible to obtain an ordinal scale from questionnaires. One of the most common, if not the most common is the multiple-choice test which has the choices: extremely likely/agreeable, likely/agreeable, neutral, unlikely/disagreeable, and extremely/very unlikely/disagreeable.

**Nominal scale:** The lowest level of measurement and the simplest in science is that of classification. In classifying we attempt to sort elements into categories with respect to a particular attribute. This is the nominal scale. On this scale we can only say if some element possesses a particular attribute but cannot even rank them according to some scale on a hierarchy based on the intensity of possession of that attribute. We can only think of creating sets based on the possession of some property and apply the operations for sets. In this sense the set operations are the most primitive of operations of mathematics. It ranks so low on the scale or hierarchy that we all instinctively do it. Even an animal that can tell food from nonfood can be said to have learned or can be said to know about set operations instinctively. Whatever kind of logic that flows from this must obviously be related to set theory in some way. If it were not so, how would animals be able to do it?

Consider now a simple question: Is 100°F twice as hot as 50°F? In a quick and unscientific poll via email in my GER class the students split approximately 50-50 on this question. There are similar questions that can produce similar splits: What is the volume of stereo player? What is the average pitch of voice of a female? Is it twice as high as that of a male? What is different about these than say measuring distance or weight? The simplest reason is that we are not familiar with the scales used in measuring them. The second reason is that normally we are used to interval scales and not absolute scales say in measuring distance. Let’s explore the implications of the differences and the difficulties using the temperature scale. Heat transfer rate is usually approximated as linearly proportional to the temperature difference between the two bodies

\[ q' = \alpha(T_1 - T_2) \]  
(A.1a)

and here it doesn’t make any difference whether we use the absolute (i.e., Kelvin or Rankine) or the relative (Fahrenheit or Celsius) temperature scale. However, the efficiency of a heat engine is proportional to the ratio of the two temperatures i.e.

\[ h = 1 - (T_1/T_2) = (T_2 - T_1)/T_2 \]  
(A.1b)

so that the absolute scale must be used. In order to understand clearly what this means, it is instructive to look into the history of development of the temperature scale. It was noticed that the level of some fluid put in a glass tube was correlated with the ambient temperature. Fahrenheit decided to put some numbers to this effect. In order to normalize it, he used the body temperature as a constant and marked that as 100°F. He then arbitrarily put some tick marks on the thermometer and on this scale we got water to freeze at 32°F and to boil at 212°F. In order to put it into more logical/scientific form Celsius decided to use the freezing and boiling point of water and marked these off as 0°C and 100°C. However, this wasn’t enough. Because of theoretical considerations from thermodynamics Lord Kelvin realized that an absolute scale was needed and hence we now have the absolute scale in which water freezes at 273 K and boils at 373 K, so that absolute zero is at -273°C. In the absolute Fahrenheit scale also called the Rankine scale, absolute zero occurs at about -460°F. As a testimony to the insight of Kelvin we should note that no one to date has been able to produce any temperature colder than absolute zero, that is less than -460°F. Hence we can see the answer to our question since the ratio is (460 + 100)/(460 + 50) = 500/510 = 1.068 so that 100°F is not twice as hot as 50°F (Figure A4).

We should think of intelligence as being more than simply what these tests purport to measure. It’s a process and the tests measure the end result of some computation or process. If we think about intelligence in a more general sense, as in animals, we see that survival techniques
are a big part of it. Animals adapt to the environment; so do humans. Humans adapt to slavery as well as to being masters; they adapt to the slums as well as to the suburbia. Deceit is part of intelligence according to some since an artificially intelligent entity (AIE) wouldn’t pass the Turing Test (TT) if it only told the truth. This we certainly don’t seem to measure directly in IQ tests. One of the highly stylized facts is that intelligence should be close to 100 per cent inherited since if it were not true we should be teaching statistics and calculus to chimps and dogs. And yet intelligence as a concept has to take into account both machine intelligence (MI) and intelligence of animals. The real question is if correlation analysis, regression analysis or factor analysis does what it’s supposed to do?

In comparing intelligences of humans without an absolute intelligence scale common sense says that ratios should be used and not differences. And indeed they are used. The tests are structured so that the results are Gaussian and that the average is 100. However this is an interval [i.e. relative] scale and not an absolute scale. On this test a difference of 15 points really boils down to 15/100 or 15 per cent although this is not explicitly stated either because it’s too obvious or because of the preoccupation with making things scale free. But we lose information every time we reduce sets of numbers to a single number and we lose some information when we rescale the original variables. If intelligence is something that all living entities [and possibly even machines] possess because our basic concepts of intelligence dictate that some kind of intelligence is possessed by animals, then it has to be measured on an absolute intelligence scale. Now suppose we add some constant to the scores on the tests to make sure we are measuring them on an absolute intelligence scale, say, about 7,000,000. Then the average score for whites is 7,000,100 and for blacks 7,000,085. Now, what person in his right mind would claim that the difference (7,000,100 - 7,000,085)/7,000,100 = 2.142910^-6 = 0.0000021429 which is 0.00021429 per cent has any kind of significance? In fact, it doesn’t. And it is also completely consistent with the kinds of results we should expect. Almost all of our intelligence is inherited, and we should all score very high when compared against say dogs, salamanders, or even chimps.

Appendix A.2: Potential and its realization
Other types of potential models, with yet another way of creating realization from potential can be created. Two examples will be given to show the futility of the standard arguments against intelligence tests. For one thing, it’s easy to define some function say

\[ \Psi = \frac{1}{2}(eE^2 + \alpha A^2 + \alpha S^2) \]  

(A2.1)

<table>
<thead>
<tr>
<th>Fahrenheit</th>
<th>Celsius</th>
<th>Kelvin</th>
</tr>
</thead>
<tbody>
<tr>
<td>212</td>
<td>100</td>
<td>373</td>
</tr>
<tr>
<td>32</td>
<td>0</td>
<td>273</td>
</tr>
</tbody>
</table>

Figure A4. The temperature scales: In physics the absolute temperature scale (Kelvin scale) must be used in order to be meaningful in equations involving ratios.
Meanwhile it is of interest to note that the problem of equalizing intervals gets a bigger play from
the mental testers and other psychometricians than it does from the sensory psychologists. In
psychometrics the need for equal units is distressingly acute, because the paraphernalia of
metric statistics (means, standard deviations, coefficients of correlation, etc.) seem to be essential
tools of the trade. The assessor of human abilities is usually knee-deep in statistical problems to
which most statistics do not apply unless his units can be equalized. Out of this quandary he
hoists himself by an act of faith in the "normal distribution" of nature's errors. If this faith is
firmly founded-in truth it is legitimate to sue the distribution of scores as a criterion for the sizes
of units—the the equalization of units is possible. It is certainly not unreasonable to believe that
his faith is often justified. What huants us is the difficuty of knowing when it isn't. Stevens
(1951:39)

where for sake of argument \( E = \) endurance, \( A = \) agility and \( S = \) strength or power. Let's name
\( Y \) the Fizikal Fitness Potential (FFP). Can we now derive a scalar (a single number) called
Physicality Quotient (PQ) or a Fizikalness Vector (FV) from this potential? Can we not derive a
single number for something having to do with physical ability, physicality, or something along
these lines? What about the scoring system of the decathlon? Does it not produce a single number
for the various events in the contests? Is it all part of what we think is about physical fitness?
There are two simple answers to these questions: PQ, or FPP or FV) is what the tests measure in
other words if the word physical was not defined in FFP, PQ or FV, well, then it is now via the
scores on these tests. So what we have done is to prescribe an operational procedure which can be
duplicated (as closely as possible) and this operational procedure is what defines whatever it is
that we were trying to define. It's an idea due to Bridgeman and comes from physics when the
physicists had similar problems with what an electron was or if it really existed. If, on the other
hand, we think that physical fitness or physical ability was something that we all know and that
they were already defined then we'd fight over whether PQ or FPP or FV measure it. The real
problem is then, is this what it's supposed to be? The whole idea of using words like potential and
actual are fraught with dangers so we should be careful. After these standard disclaimers about
problems which will probably plague this field for a long time, if not forever, we can proceed by
assuming that we can measure these variables \((E,A,S)\) via some suitable tests and that these
tests will have met some of the basic criteria required. We can then easily produce a vector from
this potential

\[
p = \text{Grad}(\Psi) = \nabla \Psi = \mathbf{u}_E eE + \mathbf{u}_A aA + \mathbf{u}_S sS = \begin{bmatrix}
    eE \\
    aA \\
    sS
\end{bmatrix} \begin{bmatrix}
    0 \\
    1 \\
    1
\end{bmatrix} = \begin{bmatrix}
    eE \\
    aA \\
    sS
\end{bmatrix}
\]

(A2.2)

where the \( \mathbf{u} \)'s are the unit vectors in the proper directions. For what follows bold lower case Latin
letters will always represent vectors or tensors of rank 1, bold upper case Latin letters matrices or
tensors of rank 2, and normal Latin or Greek letters will be scalars. The letters \( i, j, k, n \) will be
used for integers. For what reason we'd do this is not important for now. If we want now to
produce a single scalar number from this vector we can use some general norm such as in
Equation (4)

\[
\rho = |p| = (eE^2 + aA^2 + sS^2)^{(1/n)}
\]

(A2.3a)

For \( e = a = \sigma = i = f = k = 1 \) and \( n = 2 \) this becomes the standard Euclidean norm. Note that
we want even powers just in case the numbers we measure may be negative assuming that the
components of the vector are "positive" quantities so that we don't want some components of the
vectors diminishing the norm of the vector. Note that we could have used other components such
as height, weight, etc. and some of these could be very important in predicting success in sports
such as basketball or football. And we can produce many kinds of measures from something of more complicated form such as

$$\rho = |p| = (E^* A^t S^t) + (E^* A^* S) + \alpha A^\delta + \sigma S^{3k}) \gamma/h$$  \hspace{1cm} (A2.3b)

from which we can produce measures such as EAS+ EASE(E+A+S) or as EAS +E+A+S or EASE(E+A+S). Since product terms are useful in equalizing the various components in the single scalar measure, the latter could be used to further weight the weighted sums. The real question is if we have some fundamental reasoning behind selecting any of these or even selecting some other versions using logarithms. The scalar measure (i.e. the single number) can be changed if any negative factors are being used as components of the vector. For example, one does not normally find tall persons in gymnastics but lots of them in basketball. So we can easily work out different scales for different sports. The real question is what weights should be given for the FFQ or PQ and what the number really means. Assuming that we don't have much data we can always use these numbers in statistical studies judging the success of the subjects in various sports. Other possibilities exist and come to mind. For example, we could have used cubic powers for the potential $Y$, then the vector would have had squares of the variables. We could have then calculated the divergence of the vector and obtained another scalar.

$$\Phi = \nabla \cdot \nabla \Psi = \nabla^2 \Psi = e + \alpha + \sigma = \text{constant}$$ \hspace{1cm} (A2.4)

But what could it mean? In this case we are at a loss, at least momentarily, to know what these single scalar numbers could mean. However it is not in general true that a single number has no interpretation or that it cannot be done, although in general and for results such as for CA, PQ or FFQ, we can and probably will always have arguments as to the appropriateness of the given scalar.

If it is to have any enduring value, sound debunking must do more than replace one social prejudice with another. It must use more adequate biology to drive out fallacious ideas. S. Jay Gould (1981: 322)

In any case, the most important thing is whether they correlate with things in the real world. And in extending the analogy to the test of intelligence we have at least two related problems. One is that they do correlate with success in our society, and not too many people would or should argue with this. It measures something which society seems to value. It is the other problem that's caused the hullabaloo. Is it possible that these scores mean something about the genetic characteristics of the subjects. And here we must get more technical. Suppose for simplification we decompose cognitive capability or ability into usual Verbal (V[serial] and Spatial(S)[parallel]) (ignoring others such as Mathematical, Logical or Symbolic). We could, of course, use other fundamental criteria such as Serial or Parallel processing capability. We realize, of course, that there is no such neat division into any of these categories. Story telling and following takes place in at least 4D (3 space dimensions which are conceived in parallel and one time dimension, which is Serial). Similarly problem solving involves 4D. Mathematical talent/knowledge consists of solving 4D problems in detail and precision. However suppose that we want to derive a single number from these scores. Assuming that there's some kind of a maximum capacity for the human brain (and there is since it's infinite) we must have something like $V+S=k=\text{constant}$. We can define the potential as

$$\Phi = V^m S^n.$$ \hspace{1cm} (A2.5)

Then the actuality vector using the same ideas would be

$$\alpha = \nabla \Phi = V^m V^{-1} S^t \hat{e}_v + S^m S^{-1} \hat{e}_s = \Phi(\alpha/V) \hat{e}_\nu + (S/S) \hat{e}_s = VS$$ \hspace{1cm} (A2.6)
\[ \nabla \nabla F = u(u - 1) y^{x-2} S^x + s(s - 1) y^x S^{y-2} \quad (A2.7) \]

Evidently for this 'model' we'd need \( s, x > 2 \) in order for some meaning to be attached to these in the standard sense. There could be good reasons for choosing a multiplicative version of computing a number. If we want to give equal weight to being rounded in all areas, we'd do well to use a multiplicative number since it's maximized if the components are about equal, and since a serious deficiency in one component would seem to have a large negative impact on observed/hypothesized intelligence we might do such a thing. For example, to maximize the surface area of a rectangle for a given perimeter we can see that we have to maximize \( A = hw \) where \( 2h + 2w = P \) constant. Therefore since we have \( h = (P - 2w)/2 \) we can substitute into \( A \) and differentiate \( A \) wrt \( h \) to obtain the result that the maximum occurs at \( h = w \). So we can add a multiplicative component to the measurements of various components to derive a single number. If anything it is the lack of such measurements that are a greater hindrance to the resolution of problems in this field and not the existence of tests.

Another example of 'potential' and its realization is how people choose to "expend" (even if they do this slowly over time) their intellectual energy in the dichotomy between the generalist and the expert. It is said that the expert knows everything about nothing. In the same way the generalist knows nothing about everything. These can be represented as in Figure A5.

Appendix A.3: Exact differentials and path functions/integrals

The distinction between the related concepts state and process is an important one. There are mathematical definitions and consequences of these ideas. A state (or property) is a point function. The state of any system is the values of its state vector (a bundle of properties which characterizes a system). If we use these variables as coordinates then any state of the system is a point in this n-dimensional space of properties/characteristics. Conversely each state of the system can be represented by a single point on the diagram (of this space). For example for an ideal gas the state variables are temperature, pressure, volume, etc. Each color can be represented as a point in the 3-D space spanned by the R, G and B vectors. Intelligence is commonly accepted to be a state variable, i.e. a point. The scalar, Spearman's \( g \), (single number, not a vector) can be obtained from this vector by using a distance metric. The argument that the values of the components cannot be obtained from the scalar, \( g \), may be valid depending on the distance metric (however, the distance metric may be devised in a way in which the components can be obtained from the scalar). Distance on a metric space is a function only of the end points i.e. between two states. However, the determination of some quantities requires more than the knowledge simply of the end states but requires a specification of a particular path between these points. These are called path functions. The commonest example of a path function is the length.

Figure A5.
The potential and the expert: In (a) the expert has more general knowledge than the expert in (b). The isoclines are types of knowledge.
of a curve. Another example is the work done by an expanding gas is a path function. So is the heat (transferred). In that sense work and heat are interactions between systems, not characteristics of systems. Intuitively, when we talk about small changes or small quantities we use the differentials \( \delta x \) or \( \delta x \). When we want to sum up an infinite number of these, and use integration we use the notation \( \int \delta P \) for an infinitesimal amount of or \( \int \delta Q \), for an infinitesimal quantity of heat or \( \int \delta s \), for an infinitesimal length of a curve. However the crucial difference is that although there may exist a function \( f(x) = \frac{dP}{dx} \) so that

\[
\int_a^b dF = \int_a^b f(x) dx = F(x)|_a^b = F(b) - F(a)
\] (A3.1)

there is no function \( Q \), such that

\[
\int_a^b \delta Q = Q(x)|_a^b = Q(b) - Q(a)
\] (A3.2)

Instead we write \( \int_a^b \delta Q = Q_{ab} \), meaning that \( Q_{ab} \) is the quantity of heat transferred during the process from point \( a \) to point \( b \). Similarly because the infinitesimal length of a curve in the plane is given by \( ds = \sqrt{dy^2 + dx^2} \) we cannot integrate \( ds \) and obtain

\[
S(b) - S(a) = \int_a^b ds = \int_a^b dx
\] (A3.3)

but instead first the curve \( y = f(x) \) must be specified so that we can use compute \( dy/dx \) and then use

\[
ds = dx\sqrt{1 + \left(\frac{dy}{dx}\right)^2}
\] (A3.4)

Equivalently, if \( z \) is a function of two independent variables \( x \) and \( y \), and this relationship is given by \( z = f(x, y) \) then \( z \) is a point function. The differential \( dz \) of a point function is an exact differential and given by

\[
dz = \left(\frac{\partial z}{\partial x}\right)dx + \left(\frac{\partial z}{\partial y}\right)dy
\] (A3.5)

Consequently if a differential of form

\[
dz = Mdx + Ndy
\] (A3.6)

is given, it is an exact differential only if

\[
\frac{\partial M}{\partial y} = \frac{\partial N}{\partial x}
\] (A3.7)

Therefore in the mathematical function used for the simple two-factor (nature-nurture) Intelligence Function (Appendix A.4), the environmental path taken does make a difference in the final result which is assumed to be a state function (although computed from mental processes). As a very simple example of the kinds of dynamics we can expect we can try a simple path of increasing richness of environment by first writing the equation for the increase in intelligence as
in which we now have used \( E \) to show that the intelligence level may be modeled as approaching some constant value determined by the environment since when \( \gamma = aE \) the derivative is zero and hence the intelligence has hit its maximum level (steady state) and will increase no longer. The solution of this is quite easily shown to be

\[
I(t) = aE - Ce^{-\gamma t}
\]

where \( C \) is an arbitrary parameter. As time increases the second term goes to zero so that the maximum intelligence (a simple function proportional to the environmental level) is reached. It is already known that IQ scores have been increasing for about half a century and that the levels of education, nutrition and the general standard of living are usually thought to be the cause of it. Taking a cue from this fact and the equation above we can change the model to a more general form in which environmental richness is not constant

\[
\frac{dI}{dt} = \gamma(aE(t) - I(t))
\]

The general solution of this is

\[
I(t) = \gamma e^{-\gamma t} \int E(t)e^{\gamma t} dt + Ce^{-\gamma t}
\]

We note that this is of form Equation (24) and it came from very reasonable modeling of the brain/body growth. For short periods of time we can approximate the change in the environment as a linear function of the form \( E(t) = A + Bt \) where the \( A \) is the environment to which the child was born and the B coefficient determines the linear enrichment rate of the environment. The solution is

\[
\gamma(t) = \frac{1}{\gamma}(A + Bt - Ce^{-\gamma t} - B)
\]

The plots of the increase in intelligence of three children for various values of the parameters is shown. One is born to a low SE class (\( A = 150 \)) but his environment is changing more rapidly (\( B = 0.45 \)). From the earlier equations we can try to give a meaning to the constant \( \gamma \); it simply controls the rate at which the child adapts so if anything it could be some measure of innate ability, however we have not explicitly chosen any genetic component in the equation. The form of the solution implies that this coefficient \( g \) has to do with what many would consider to be a genetic factor which controls the speed at which people learn new things and adapt to the environment. The environment then acts as a motivation factor in which each child realizes his/her potential. In the growth equation this coefficient controls the rate at which the child would reach his/her adult size. Therefore we assume that it could contain some "innate" factor.

Thus for this child the constant \( \gamma = 0.15 \). The other has started off in a relatively advantaged position (\( A = 170 \)) and is in a state where his environment is not being enriched as rapidly as the first (\( B = 0.34 \)). Furthermore \( \gamma = 0.11 \). For both we have chosen the arbitrary constant of integration \( C = 100 \). As can be seen from the plots, although the first child, if anything, is of greater innate ability (given the meaning attached to the coefficient \( \gamma \)), he cannot reach the intelligence level of the other; the difference drops from a high of about 20 down to approximately 8 at around age 8 at which time the differences actually start to increase again (Figure A6).
In the second set the parameters have been changed again, but the differences in the innate learning ability i.e. \( \gamma \) are almost equal since they are 0.03 and 0.038 respectively, however the intelligence differences actually starts to increase until approximately age 12–13 and then still persist at about 15 points until age 30. One quick lesson we can see from this model is that units of measurement or scales do matter. The coefficients were selected arbitrarily in order to produce numbers in the ranges in which many tests work i.e. IQ tests around 100, and various achievement tests which are scored between 200 and 800. These are exactly what kinds of results that we should expect because of the evidence that early experiences are important, and that what we are measuring (purportedly a function of genetic ability and not learning) is a function of the environmental path. As simple as they are these models are still much better than mindlessly regressing of everything against everything else, and then attempting to draw deep scientific conclusions from them. We don't know exactly how the increase in the environment goes so that instead of linear it could be exponential or logistic/sigmoidal. One can easily change
the environmental effects so that they are of sigmoidal shape. Many such differential equations with their solutions can be found in Hubey (1987, 1991a) related to software engineering. As an example of a function which weights early effects of the environment more heavily than the later ones, let us choose a simple exponential weighting

$$w(t) = e^{-\alpha t}$$  \hspace{1cm} (A3.13)

In addition let us choose an increasing environmental richness path (normalized single variable). A computation of an accumulation (of any behavioral observable, for example intelligence) over paths $j$ with the weighting as shown above, would be

$$X(t) = \int_0^t E(t)e^{-\alpha t} \, dt = E \int_0^t (1 - e^{-\beta t})e^{-\alpha t} \, dt \quad (A3.1e)$$
in which \( B(E) \) is of form \( \alpha E^k G^l (1 - e^{-\lambda E^m G^n}) \) with \( \delta_i = \lambda E^m G^n \) and \( E_i = \alpha E^k G^l \). These can be integrated with different values of \( \delta_i \) and \( E_i \) for a simulation of and comparison of the effects of the early environment. The early problem solving techniques learned add appreciably to later scores which was shown from the perspective of memory utilization and access time in section 2.4 More examples of this can be found in Appendix A.4.

Appendix A.4: The intelligence function constraints

The Intelligence Function of form

\[
I = \alpha E^k G^l = F(E, G)
\]  

(A4.1)

was selected only because it was the simplest such form that satisfies the conditions we expect such functions to satisfy. For \( E > 0 \) we expect positive and diminishing marginal returns

\[
F_E = \frac{\partial F}{\partial E} > 0, \quad \text{and} \quad F_{EE} = \frac{\partial^2 F}{\partial E^2} < 0
\]  

(A4.2a)

That is, we expect increases in intelligence scores with increasingly rich environments. The IQ scores of children over the century has been increasing in all the developed countries. But we expect the marginal increases to slow down as we enrich the environment because there is a top limit on the capacity of the human brain. Similarly for \( G > 0 \) we expect

\[
F_G = \frac{\partial F}{\partial G} > 0
\]  

(A4.2b)

but it is not yet clear if

\[
F_{GG} = \frac{\partial^2 F}{\partial G^2} < 0
\]  

(A4.2c)

because we do not yet understand enough about genetics. However, it seems a reasonable enough assumption for a simple beginning. At the same time we can see that if we let \( h = 1 - e \), then all the conditions of Equations (4.2) are satisfied if \( h < 1 \). In addition we can also expect

\[
\lim_{E \to 0} F_E = \lim_{G \to 0} F_G = \infty
\]  

(A4.3)

\[
\lim_{E \to -\infty} F_E = \lim_{G \to +\infty} F_G = 0
\]  

(A4.4)

if Equation (A4.2c) is true. These are the time-independent restrictions and conditions on the Intelligence Function.

As is clear the models (both the Intelligence Function of this appendix and the dynamic model in the body) still do not account explicitly for the variation in utility of early learning vs late learning. One would still have to weight early learning (i.e. environmental variable) more heavily than late learning. One such exponential weighting scheme is given as an example further below.

Since heredity changes at glacial speeds for a species we do not need to bother making \( G \) as a function of time, however in the equation above, although time is explicit to account for increasing intelligence over time, there is no time in the environmental variable itself and there is no time-dependent weighting of the environmental variable. Since we want to be able to account for the cumulative effect of the change in the environmental variable we might have to change it to

\[
I = \alpha E^k G^l \left( 1 - e^{-\lambda E^m G^n} \right) \]  

(A4.5)

Evolution of intelligence

415
\[ I = \alpha g^h \int_0^t E^s(t') \ln (1 - e^{-\lambda t'}) dt' \]  \hspace{1cm} (A4.6) \\

or even

\[ I = \alpha g^h \int_0^t E^s(t') \left( 1 - e^{-\lambda t} \int_0^{t'} E^u(u') du' \right) dt' \]  \hspace{1cm} (A4.7)

Since it has already been argued at length (everywhere and here) that what is accumulated happens across time so that the rich-get-richer scenario (the Matthew effect of education) (of the simple kind) is of form

\[ I(t) - \lambda \int_a^t K(t, \tau) I(\tau) d\tau = f(E(t), G(t), t) \quad t \in [a, b] \]  \hspace{1cm} (A4.8)

where the \( f(t) \) = intelligence, and it is the past [multiplicative] interaction of intelligence with both hereditary and environmental factors summed up over time (the past) and driven by gain a [multiplicative] interaction of environmental/social and hereditary/genetic factors. Among these interactions are the elegant integrated theory of both the Power Law of Practice (i.e., the more you practice the faster you get, or the Matthew effect), and the Fan Effect (i.e., the more you know the slower you go) (Schrager et al., 1988). Integral equations of the type

\[ I(t) - \lambda \int_a^t K(t, \tau) I(\tau) d\tau = f(t) \quad t \in [a, b] \]  \hspace{1cm} (A4.9)

are called Fredholm equations. Volterra equation of the second kind is a special case of the Fredholm equation (Cushing, 1975 or Tricomi, 1985). We can try normalizing the solution given in Equation (31) for convenience as

\[ I(t) = \frac{cH(t - \tau)(1 - e^{-\mu(t-\tau)}) + F}{C + F} \]  \hspace{1cm} (A4.10)

where \( C = \) ceiling, \( F = \) floor, \( H(t) = \) Heaviside Unit Step Function and \( 0 \leq I(t) \leq 1. \) This is to be realistic since nobody is born in any society to "zero environment", but there is a floor (which changes from society to society), and there is a practical ceiling which no matter how rich the person is cannot be exceeded.

Some typical ones are plotted below.

We can see that

\[ I_{early}(0) = \frac{F}{C + F} \quad \text{and} \quad I_{late}(0) = \frac{C(1 - e^\mu) + F}{C + F} \]  \hspace{1cm} (A4.11)

\[ I_{early}(\infty) = 1 \quad \text{and} \quad I_{late}(\infty) = 1 \]  \hspace{1cm} (A4.12)

It would not be difficult to use nonlinear equations, for example, the logistics equation (Verhulst equation) instead of these linear equations. Simple versions of the Verhulst equation is used in Rasch testing (1980). Mathematical modeling of forced binary discrimination is naturally explained via the Verhulst/logistic equation (Hubey, 1994).
Appendix A.5: Statistical testing methods
If we have a 1st order linear ODE like this
\[
\frac{dy(t)}{dt} + by(t) = f(t)
\]  
(A45.1)
we can try several easy ways to test it.

\textit{Method 1:} Take the Laplace transform and turn it into the algebraic domain obtaining
\[
sy(s) + sy(0) + by(s) = f(s)
\]  
(A5.2)

\textit{Method 2:} Transform the solution. The convolution integral becomes a product in the transform domain, so for the steady-state part we have
\[
y(s) = g(s)f(s)
\]  
(A5.3)

\textit{Method 3:} Start with the solution, the steady-state of which is of form:
\[
y(t) = \int_{0}^{t} g(t - \tau)f(\tau)d\tau
\]  
(A5.4)
Discretization of this yields a time series in which the rhs is a summation of the type
\[
\sum_{j=1}^{n} b_{j-k}f_{j}
\]  
(A5.5)
This is a simple version of the ARMA/ARIMA model. It can be done for every linear differential equation since the solution of every linear DE is in terms of a convolution integral with the kernel as the Green's function. However, there are even more apt analogs in which the concept of linearity can be used to create more complex outputs and thus simulate richer class of behaviors. For example, let \( L(t) \) be a linear differential operator, then
\[
L(t)x(t) = f(t)
\]  
(A5.6)
is what we'd call an inhomogeneous linear differential equation. As earlier \( x(t) \) is a vector of output/behavior/response and \( f(t) \) is external forcing i.e. inputs/stimulus. In this sense \( L(t) \) is an

\[ I_{\text{early}}(t) = \frac{C(1-e^{-w(t)})+F}{C+F} \]
\[ I_{\text{late}}(t) = \frac{H(t-\tau)C(1-e^{-w(t-\tau)})+F}{C+F} \]

\( \tau \)
\( t \)

Figure A.8.
The effect of timing delay on environmental variables: the rich environment early in life leads to greater intelligence scores.
operator which is a description of the system being simulated. Since if we had a simple first-difference equation in $x_n$ with forcing $f_n$ as

$$x_{n+1} = x_n + a f_n$$  \hspace{1cm} (A5.7)

the solution is

$$x_{n+1} = x_0 + a \sum_{n=0}^n a^n f_n$$  \hspace{1cm} (A5.8)

therefore the input into the system (i.e. stimulus) completely determines the response of the system. What happens if we try this idea on coupled equations? As given even if $L(t)$ consists of constant coefficients the behavior is intimately related to the linear vector systems such as $y = Ax$ since it can be approximated discretely as shown below.

$$\frac{dx}{dt} = ay$$  \hspace{1cm} (A5.9a)

$$\frac{dy}{dt} = -bx$$  \hspace{1cm} (A5.9b)

can be written as

$$x_{n+1} = x_n + ay_n$$  \hspace{1cm} (A5.10a)

$$y_{n+1} = y_n + bx_n$$  \hspace{1cm} (A5.10b)

which can be written as the vector equation

$$z_{n+1} = Az_n$$  \hspace{1cm} (A5.11)

We can, of course, add external forcing into the system by writing it as

$$z_{n+1} = Az_n + f_n$$  \hspace{1cm} (A5.12)

For the differential equation formulation we can differentiate the first (A7.11) and substitute the second (A7.11) to obtain

$$\frac{d^2x}{dt^2} + abx = f(t)$$  \hspace{1cm} (A5.13)

where we have added the forcing $f(t)$ to make it similar to the discrete version. This can also be written in the formalism of linear theory since the integral part is the inverse of the linear differential operator. Therefore we can see the same principles in operation in both the differential equation and linear algebra case. In order to show the correspondence more clearly we can eliminate $y(n)$ from the discrete equations to obtain

$$x_{n+1} = 2x_n - (1 + b)x_{n-1}$$  \hspace{1cm} (A5.14)

which is a second order difference equation corresponding to the second order differential Equation (A5.13). Without developing more comprehensive theory we can show that both $x_n$ and
\( y_n \) depend on their past because of the coupling. For example, simply iterating and substituting into (A5.11) we can see that

\[
x_{n+1} = x_0 + a \sum_{n=0}^{\infty} y_n
\]  
(A5.15a)

Similarly

\[
y_{n+1} = y_0 + a \sum_{n=0}^{\infty} x_n
\]  
(A5.15b)

It is clear that such simple linear equations cannot model HEPSIs (Historico Econo Politico Social Intelligent) since there is no learning in which the system characteristics changes in response to the inputs (stimulus). In light of the fact that we can write any order difference equation as a system of difference equations of the type then we can use the same iteration technique to obtain the solution to a multidimensional (vector) difference equation with the difference that we must iterate a vector equation

\[
z_{n+1} = A^{n+1} z_{n-j} + \sum_l A^j z_{n-j}
\]  
(A5.16a)

substituting \( j = n \) we obtain

\[
z_{n+1} = A^{n+1} z_{0} + \sum_l A^j z_{n-j}
\]  
(A5.16b)

We see that the solution at any time is determined by the weighted average of input over time (modified by the characteristics of the system itself as exemplified by the matrix \( A \)). The initial state \( z(0) \) of the system is propagated forward by the system characteristics matrix \( A \) which is modified by the sum of the forcing (input or stimulus). However, as can be easily seen, the basic characteristics of the system as manifested in \( A \) cannot and does not change. In other words this formalism cannot model/simulate a self-learning system. It models whose behavior or response to inputs is time invariant such as that of a spring or a string on a guitar. A string on a guitar does not learn anything from the last time it was struck, and does not change its behavior in response to it ideally speaking. Practically speaking, it might stretch, the stringing might loosen etc. but in physical analysis we ignore such small changes. Indeed, if the string is stretched over its elastic limit, its behavior is not linear at all but again we have taken the simplest idealized behavior.

We should be careful here to mention the caveat that much of the rich behavior of such recurrence relations is obtained from the nonlinearity as is well known. And since many of these nonlinear systems are deterministic but chaotic we can already see the germ of the solution of the philosophers' conundrum of determinism vs randomness. We do not want life to be deterministic since it implies fatalism but we cannot assert that everything is random since it would contradict the everyday real observation that much of what we do is purposive. We do not behave at random and we'd like to think that our actions are results rational(s) beings would make repeatedly under exactly the same conditions. Since we can never have exactly the same conditions (because of learning) we cannot use the basic methodology of the hard sciences in which we repeat experiments in a laboratory. At the same time we cannot be predictable for that would make us robots or automatons. What we want is to have both determinism and unpredictability. The recent advances in mathematics is exciting because we can have both rational/logical action and at the same time uncertainty of behavior. In one case, we have fuzzy logic in which we can use uncertainty not due to chance or probability which is a better
philosophical idea. In the other case we have chaos in which we have deterministic algorithm but unpredictable output. In a sense, stochastic or random is infinite dimensional chaos implying figuratively that there are degrees of unpredictability which is the embedding dimension.

However the differential equation formulation allows explanations of the type in which many of the criticisms made against the previous models are not valid. The main reasons why such models are not used is probably because of the philosophical attitude of those who work in the HEPSI systems field. The equations above were ordinary, linear, deterministic, forced (and homogenous) differential equations with constant coefficients (scalar and vector types). The main thrust of this appendix is to show how and why what is needed or is thought to be needed to model/simulate HEPSI systems is or can be shown to be in differential equations. Naturally, those equations that are not solvable can always be solved numerically (i.e. simulated on a digital computer) to (almost) any degree of accuracy. It is possible that some of the complex behavior shown in numerical solutions of nonlinear differential equations is due to the inherent problems of error propagation in discretization (which is a form of quantization) noise.

Appendix A.6: Absolute measurement scales and fuzzy-logic, heritability-multiplicativity paradoxes
If the additive model of intelligence is used then we can use the method of virtual variation to analyze how the variations in environment and genetic makeup will influence intelligence (scores) by computing the total differential (or variation) as

\[ dI = v \cdot dE + h \cdot dG \]  
(A6.1)

The variation above, say in \( E \), is amplified through the effect of the coefficient \( v \), and thus the share of the total variation in \( I \), is due to \( v \cdot dE \). However since the variation could be both positive or negative, to obtain the magnitude of the variation, or a number with essentially the same dimensional significance as the variance we should compute

\[ dI^2 = v^2 \cdot dE^2 + h^2 \cdot dG^2 + 2hv \cdot dE \cdot dG \]  
(A6.2)

For the nonlinear case explicitly considered in above the variation is

\[ dI = (aeE^{G^{-1}}G^b) \cdot dE + (ahE^{G^{-1}}G^b) \cdot dG \]  
(A6.3)

But now the total variation due to the variation of \( E \), is amplified via the effect of coefficients which obviously have genetic components. Can we now compute \( dI^2 \) and still claim that the coefficient of \( dE^2 \) is still only the environment effect? Clearly no. The type of model does make a significant difference in the interpretation of results. Continuing with the computations, since the terms \( dE^2 \) and \( dG^2 \) are the virtual variations which are the equivalents of variances in statistics and since we explain the total variance statistically via \( V = V_E + V_G + V_{CE} \) the likely conclusion is that the statistical methods work on variables measured on the absolute (ratio) scale and furthermore on some "natural scale". For the linear case we could not make a claim (unless the absolute scale were used i.e. \( E = G = 1 \)) that the coefficient of \( dE \) was the variation due only to the change in the environment. For the nonlinear case, the change in genetic endowment is amplified by environment so that we could not readily claim that the second term in Equation (A6.2) is the variation due to genes whereas in the linear case, the variation in \( G \) is merely amplified by a constant \( h \), which we assume has to do with genetics. Therefore, even analysis of variance is not scale free and it is not model free. In other words, we are faced with two conflicting views. One of them is that we should introduce new "dimensionless" variables \( E \) and \( G \), and then thus compute the variation as \( dI = dE + dG \) whereby Equation (A6.2) becomes

\[ dI^2 = dE^2 + dG^2 + 2dE \cdot dG = V = V_E + V_G + V_{CE} \]  
(A6.4)
where the new variables are the natural (absolute/ratio) scale variables. The second view is that we must have something like $e + h \leq 0.5$ since we already know that we should have $0 \leq I \leq 1$ in order to have the additive model valid as the Taylor expansion of the multiplicative model. Since we also want $0 \leq E, G \leq 1$, unless condition $e + h \leq 0.5$ is valid the intelligence measured will not be in the normalized range. We have already produced an approximately linear model in the neighborhood of human level intelligence (i.e. Equation (18)) from the basic nonlinear-multiplicative model, in which we set $E_0 = G_0 = 1$. We also noted that setting $e = 1 - h$ (Appendix A.4) fulfills some of the conditions for the Intelligence Function. Now if we let $a = 1/2$ (with the new variables we can see that Equation (18) becomes

$$I = \bar{E} + \bar{G} - 1$$

(A6.5)

Now the variations and the variances are

$$\delta I = \delta \bar{E} + \delta \bar{G}$$

and

$$\delta I^2 = \delta \bar{E}^2 + \delta \bar{G}^2 + 2\delta \bar{E} \cdot \delta \bar{G} = V = V_E + V_G + V_{EG}$$

(A6.6)

exactly as they should be. Because we have obtained the linear approximation from the nonlinear one, it is valid only in the neighborhood of $\bar{E} \approx \bar{G} \approx 1$. Certainly we must have $\bar{E} + \bar{G} > 1$ otherwise we will obtain negative intelligence. Even this condition is not stringent enough since it would imply that human genetic endowment ranks along that of viruses or that our environment is about equivalent to that of speechless beasts. In order to be valid because of the approximations inherent, we'd need the constraint $\bar{E} + \bar{G} \approx 2$. The other kinds of nonlinear formulations repeated below for convenience in slightly different form

$$I = cE \cdot e^{G^a}$$

(A6.7a)

$$e^{E \cdot G^a} = \text{cons}$$

(A6.7b)

have the virtual variations

$$\delta I = c(e^{E \cdot \delta G} \delta E + \delta E \cdot e^{G} \delta G)$$

(A6.8a)

$$\delta I = e^{E \cdot G^a} (G^a E \delta a - \delta E + a^a G^{a-1} \delta G)$$

(A6.8b)

$$\delta I = e^{E \cdot G^a} (E \cdot a^{-1} \delta E + a^{a-1} \delta G)$$

(A6.8c)

All of these are restricted to the interval $[0,1]$ in order to correctly model the environment-genetic interaction, and they are on an absolute (ratio) scale. As can be seen, the amount of "interaction" of genes and environment in Equation (A6.7b) is much greater than the others in that it is difficult to separate the variation due to one of the factors from the other. Any of these could have been used in the differential equations model of learning and development of intelligence. Only the simplest was used for convenience. The best fitting model should probably be used in studies.

There are, as can be seen, great benefits from having measurements which are on an absolute "natural" scale and in which the nonlinear interaction is explicit. One of the biggest problems, if not the biggest, that of finding natural dimensions such as Time, Length, Mass, and Temperature in the physical sciences has been overcome with the combination of statistical analysis, differential calculus, and a fuzzy-logical AND (i.e. intersection). What is even more to the point is that we now have many choices of using various different types of infinite-valued
logic (i.e. fuzzy-logic see Hubey (1998) or Klir & Yuan (1995) for examples) for modeling social phenomena of this type. As a simple example, a chimp raised in a human environment will essentially still have chimp-level intelligence, and a human raised in a chimp environment (feral children) will have chimp-level intelligence. The simplest expression of this idea is via the fuzzy-AND introduced by Zadeh (1965; 1978; 1987), which is a simple minimum function. Thus

\[ I_{\text{chimp}} = \text{Min}(E_{\text{human}}, G_{\text{chimp}}) = \text{Min}(G_{\text{chimp}}, 1) = G_{\text{chimp}} \ll 1 \quad (A6.9a) \]

\[ I_{\text{human}} = \text{Min}(E_{\text{chimp}}, G_{\text{human}}) = \text{Min}(G_{\text{chimp}}, 1) = E_{\text{chimp}} \ll 1 \quad (A6.9b) \]

Over a global scale, there is no doubt that the environment/culture created is mostly a function of genetics, and that intelligence is mostly determined by genetics, but on local scales, the environment acts essentially as a stochastic variable in the determination/realization of intelligence. There are many more such fuzzy-AND functions that could possibly be more appropriate. What is of further interest is that the multiplicative model (which is logically the correct model) produces anomalies of its own. The multiplicative model consists of a family of hyperbolic curves as shown in Figure A9. Although the first person \((E_1, G_1)\) scores higher, \(I_1\), his genetic endowment is lower than the second person \((E_2, G_2)\) who scores lower, \(I_2\), on the intelligence test.

Appendix A.7: The Meaning of nonlinearity in dynamics & learning
Starting from the perspective of dynamics there are 4 factors:

1. initial conditions
2. laws of evolution (i.e. the particular form of the DEs)
3. sources of randomness
   a. parameters
   b. forcing
   c. initial conditions
4. the need to have a model which can incorporate learning

All of these are satisfied with general nonlinear differential equations. Gelman calls (3c) "frozen accidents". The step (3b) is needed to model outside influences (i.e. convergence). Step (3a) (and nonlinearity) is needed to model learning/modification and evolution. Any set of ordinary DEs can be written in canonical form as

\[ \frac{dy}{dt} = G(y(t), f(t)) \quad (A7.1) \]

For example a simple linear and ordinary DE is

\[ \frac{d}{dt} y + a(t)y(t) = f(t) + w(t) \quad (A17.2) \]

The forcing \(f(t)\) or \(w(t)\) can be deterministic or random. If \(f(t)\) is deterministic and \(w(t)\) is random then we can think of them as external sources of influence. The particular characteristics of the system (i.e. internal characterization) is given by the left-hand-side (lhs) which is 1st
order/degree, and in this case has a time-dependent parameter $a(t)$. If $a(t)$ were constant then it too would be a "frozen" accident or parameter. Furthermore if $a(t)$ also had sources of randomness then we'd be able to represent systems that fluctuated due to randomness. Still further, the initial conditions of the system are not known, and those initial conditions will affect the outcome $y(t)$ at any time $t$, so that we have yet one more source of randomness which because of the way language started is yet another frozen accident. In this paper it will be shown that those properties/characteristics of the brain/mind that are allegedly impossible to model in any abstract way can be done quite fruitfully combining some of the understanding of the recent past dealing with nonlinear dynamics. In fact, these simple analogical models are also the type of models we need for an understanding of the basic issues in all intelligent systems including those that are ensembles of intelligent beings such as historic-econo-politico-social-intelligent (HEPSI) systems.

Linear systems
The definition of linearity is easily explained in terms of operator. Normally in mathematics we deal with a set of objects upon which another set of objects (called operators) act thus changing the objects of our attention to other objects. A linear operator $L()$ is then an operator whose behavior is describable as

$$L(x + y) = L(x) + L(y) \quad (A17.3)$$

In other words, it is an operator that commutes with addition. It is much easier to see if we write it as

$$L(A(x,y)) = A(L(x,y)) \quad (A17.4)$$

Since we normally deal with binary operators i.e. an operator that operates on only two objects, we could have written $x + y$ as $A(x,y)$ or $(x+y)$ or even as $(x,y)+$. The first notation is called infix notation and is what we normally use for addition, the second prefix and last postfix. As examples of linear operators we have the matrix operator that operates on a vector (array) which we could write as $y = Ax$. This operation is quite common in the social sciences because of correlation-regression analysis. A one-dimensional example of a dynamic behavior of a system can be written as

$$w_{n+1} = bw_n \quad (A7.5)$$

Figure A9.
Intelligence function/potential paradox: The linear relationship
$I = E + G - 1$ is approximate and valid only around $E = G = 1$, therefore the paradox is more difficult to achieve or display. If however, the nonlinear version is used, we can make use of the full scale
By substituting repeatedly we can see that the solution for all time is

$$u_n = b^n u_0$$ \hspace{1cm} (A7.6)

Since (A7.6) is a linear equation i.e. can be represented as a vector equation, it is easy to see that the solution can also be represented as a vector equation. However, this formalism is also sufficient for representing multi-dimensional dynamics since we can simply write

$$x_{n+1} = Ax_n$$ \hspace{1cm} (A7.7)

where the evolution of vector \( x \) (at time \( n + 1 \)) i.e. \( x(n + 1) \) is determined by matrix \( A \). This can be taken to be a very simple model of an input/output system. It is used in economics, however it can also be used as a toy mathematical model of simple behaviorism in the form \( r(t) = A(t)x(t) \) in which \( r(t) \) is the response at time \( t \), to a stimulus (input) \( s(t) \) at time \( t \). We can allow the characteristics of the system represented by \( A(t) \) to be a function of time to allow for slightly more flexibility that is normally allowed in simplistic behaviorism. We cannot expect intelligent (learning) systems to respond to the same stimulus in exactly the same way each time. The simple behaviorist view which was meant for low intelligence organisms or for particular aspects of the behavior of HEPSIs such as operant conditioning or classical conditioning cannot be expected to hold at such simplistic levels, although there are elements of such behavior. For example, if a lion were fed at exactly noon everyday and it refused to eat any more for five hours, we can approximate his approximate behavior by making the characteristic/evolution matrix \( A(t) \) appropriately dependent on time instead of constant. We have already discussed linear systems and shown their inadequacies. We would now like to ask the question of what more is needed to model HEPSIs. We can already see that we need a model in which the characteristics of the system changes in response to inputs (stimuli). In a differential equation, the characteristics of the system is a function of the operator. The number of derivatives, their signs, the values of the coefficients, and whether the coefficients are functions of time or are random all determine the response of the system to inputs/stimuli (or forcing/source). The differential equations model explicitly takes into account both the characteristics of the system and the environment (inputs). Secondly by providing us with the possibility of making the forcing/environment deterministic or random or a combination of both it gives us a greater freedom in modeling and provides a greater expressivity in modeling complex systems. Furthermore since it can always be replaced to discrete forms which may be thought of as cellular automata, or iterations of the type that produce fractals, or chaos, and since cellular automata are equivalent to Turing machines, it gives us the means to show that Turing machines are capable of producing a richer repertoire of behavior which can easily be sufficient to be a HEPSI system. To see this we can look at an example of a partial differential equations and how it relates to Turing machines and cellular automata. For example the elliptic partial differential equations in two dimensions known as the Poisson equations is given by

$$\frac{d^2}{dx^2} u(x, y) + \frac{d^2}{dy^2} u(x, y) = f(x, y)$$ \hspace{1cm} (A7.8)

where the values at the boundaries are fixed and given. Typically discretization of this equation involves expressing the relationship of a point \( u(i, j) \) in a grid in terms of the values which are neighbors of this grid. The central-difference method (Burden, Faires, Reynolds, 1981: 511) gives

$$2 \left( \frac{h}{k} \right)^2 u_{i, j} - u_{i+1, j} + u_{i-1, j} + \left( \frac{h}{k} \right)^2 u_{i, j+1} + u_{i, j-1} = -h^2 f_{x, y}$$ \hspace{1cm} (A7.9)

where each \( u(i,j) \) is an approximation for \( u(x, y) \) which is the discrete grid representation of the continuum problem. This can be re-expressed as a vector equation and solved Gauss-Seidel
method. More complex cases can be solved iteratively. Given the relationship between cellular automata and Turing machines we can see that solutions of continuum problems can be approximately solved by Turing machines. Therefore there must be more to the protests against artificial/machine intelligence than simply the analog capability of natural brains. Simple bending of a metal ruler takes much more equation and more mathematical sophistication than that. Do we expect human behavior to be simpler than a deforming ruler? A linear differential equation easily has nonlinear solutions. What this means is that we can change our domain of inquiry. So in the differential domain we can still create such models from which we can obtain nonlinear solutions. This is another attempt at simplification on the one hand (of model creation) and at the same time producing the ability to create more powerful methods which can yield complex solutions. The answer, almost as a cliche, is in nonlinearity. First of all, we should note that the nonlinearity is in the differential equation and not in the solution; that is, linear differential equations almost always have nonlinear solutions. But nonlinear differential equations are almost always not solvable in closed form. Nonlinear differential equations are those in which the dependent variable or its derivatives do not occur to any power greater than one and do not occur in multiplicative form. Almost all such models are linear in some suitable way. It is could be \( y = Ax \) where \( A \) is matrix and \( y, x \) vectors but there are lots of other ways. Statistical correlation regression tests etc. are of this type more or less. Nonlinearity cannot usually be handled because it is too difficult. It is only now with the power of the digital computer that numerical solutions can be found to some nonlinear differential & integral equations. But this is a true minimum.

**Linear DEs, and time-dependent coefficients, and the black box**

Starting with the simple linear ODE (A7.2) or (25) with \( u(t) = 0 \) and \( a(t) = a = \) constant we obtain the solution as in Equation (27). The solution is exponential and not too interesting. The meaning given to it is that \( f(t) \) is external/exogeneous and the coefficient \( a(t) \) or \( a \) is a system characteristic, which is a structural or endogeneous parameter. The steady state solution has the "black box" formulation since a discretization of the integral is essentially a matrix multiplied by \( f \) which is a vector. With the initial condition I.C. \( y(0) = 0 \) the solution is

\[
y(t - \tau) = \int f(\tau)e^{-a(t-\tau)}d\tau
\]

(A17.10)

Another example of such a system is the set of equations given by Equation (A7.5). Both coefficients are constant although the internal parameters of the system do affect each other. The easiest way to make the ODE model more powerful is to allow the coefficients to become functions of time, for example as in Equation (A7.2). But it is still linear and \( f(t) \) is still deterministic. In a "learning system" then this endogeneous parameter should somehow be affected by the inputs (over time) \( f(t) \). Suppose the equation is of type

\[
\frac{d}{dt}y + a(f(t))y(t) = f(t)
\]

(17.11)

now the system parameter is a function of the external/endogeneous variable and we can think of the system as changing its characteristics over time. In addition we know from the solution of the linear first order DE that \( f(t) \) is the endogeneous variable. We could also make that a function of the system variable itself so that there is a feedback loop created. We now have

\[
\frac{d}{dt}y + a(f(t))y(t) = f(y(t))
\]

(A7.12)
To create a simple example of this kind of a system let \( a(f(t)) = -x(t) \) and \( f(y(t)) = -y(t) \).

\[
\frac{dy}{dt} = -y + xy \tag{A17.13}
\]

Let another equation of the same type represent another system interacting with this system.

\[
\frac{dx}{dt} = x - xy \tag{A17.14}
\]

These are the Lotka-Volterra differential equations. Note the rhs of the equation for \( x \), Equation (A7.14), can be written as \( x(1 - y) \) which is \( x \) AND \( 1 - y \) in fuzzy logic. However with the coefficients such a meaning is a stretching of the imagination. The other can be written as \( y(x + 1) \) which no longer has a meaning in such fuzzy logic. That is because even in the 1st equation we can have a \((-\) \) sign which is not expressible in fuzzy logic because of the constraint of having the values in \([0,1]\). If we switch to values in, say, \([-1,1]\) we could then attach meaning to negative values. This can be done simply by shifting (i.e. define new variables). But now we still have the problem of derivatives which do not have meanings (yet) in fuzzy logic. But fuzzy logic is sort of between logic and probability theory, and if we stretch our imagination then derivatives do have meanings in probability theory (i.e. Fokker-Planck methods etc). So differential equations are more expressive and more powerful than logics.

Appendix A.8: Role of genes and heritability: a deeper peek

Many dynamic physical phenomena can be described via differential, difference or integral equations. Since most differential or integral equation cannot be solved in closed form but must be solved numerically (i.e. using a digital computer to solve for specific cases called simulation) we can also use difference equation (which are really discretized versions of differential equations). For a linear differential equation, the solution provides a mapping from the source to the output. We can think of this as a black box with the source term as the input and the solution as the output. One way in which to classify differential /is;

(a) ordinary vs. partial
(b) linear vs. nonlinear
(c) deterministic vs. random
(d) homogeneous vs. inhomogeneous
(e) constant vs. time-dependent coefficients

Biological determinism is fundamentally a theory about limits... Why should human behavioral ranges be so broad, when anatomical ranges are generally narrower?... [I] conclude that wide behavioral ranges should arise as consequences of the evolution and structural organization of the brain. Human uniqueness lies in the flexibility of what our brain can do. What is intelligence, if not the ability to face problems in an unprogrammed (or as we often say, creative) manner? Gould (1981:331)

The easier of the two has been denoted above in italics. Unfortunately, real processes in the real world happen not to be so simple. It's an fortunate fact, however, that we can approximate much phenomena of the real physical world with linear equations. But unfortunately for the biological-life-social sciences even achieving the simplest models requires competence in mathematics that transcends what is required to study physics, engineering or computer science, the so called hard
To create a simple example of this kind of a system let \( a(f(t)) = -x(t) \) and \( f(y(t)) = -y(t) \).

\[
\frac{dy}{dt} = -y + xy \tag{A17.13}
\]

Let another equation of the same type represent another system interacting with this system.

\[
\frac{dx}{dt} = x - xy \tag{A17.14}
\]

These are the Lotka-Volterra differential equations. Note the rhs of the equation for \( x \), Equation (A7.14), can be written as \( x(1 - y) \) which is \( x \) AND \( 1 - y \) in fuzzy logic. However, with the coefficients such a meaning is a stretching of the imagination. The other can be written as \( y(x + 1) \) which no longer has a meaning in such fuzzy logic. That is because even in the 1st equation we can have a \((-)\) sign which is not expressible in fuzzy logic because of the constraint of having the values in \([0,1]\). If we switch to values in, say, \([-1,1]\) we could then attach meaning to negative values. This can be done simply by shifting (i.e. define new variables). But now we still have the problem of derivatives which do not have meanings (yet) in fuzzy logic. But fuzzy logic is sort of between logic and probability theory, and if we stretch our imagination then derivatives do have meanings in probability theory (i.e. Fokker-Planck methods etc). So differential equations are more expressive and more powerful than logics.

Appendix A.8: Role of genes and heritability: a deeper peek

Many dynamic physical phenomena can be described via differential, difference or integral equations. Since most differential or integral equation cannot be solved in closed form but must be solved numerically (i.e. using a digital computer to solve for specific cases called simulation) we can also use difference equation (which are really discretized versions of differential equations). For a linear differential equation, the solution provides a mapping from the source to the output. We can think of this as a black box with the source term as the input and the solution as the output. One way in which to classify differential \( a(t) \):

(a) ordinary vs. partial
(b) linear vs. nonlinear
(c) deterministic vs. random
(d) homogeneous vs. inhomogeneous
(e) constant vs. time-dependent coefficients

*Biological determinism is fundamentally a theory about limits... Why should human behavioral ranges be so broad, when anatomical ranges are generally narrower... If we conclude that wide behavioral ranges should arise as consequences of the evolution and structural organization of the brain. Human uniqueness lies in the flexibility of what our brain can do. What is intelligence, if not the ability to face problems in an unprogrammed (or as we often say, creative) manner? Gould (1981):331*

The easier of the two has been denoted above in italics. Unfortunately, real processes in the real world happen not to be so simple. It's an fortunate fact, however, that we can approximate much phenomena of the real physical world with linear equations. But unfortunately for the biological-life-socil sciences even achieving the simplest models requires competence in mathematics that transcends what is required to study physics, engineering or computer science, the so called hard
Evolution of intelligence

sciencs. It's really the soft sciences that are hard. Because they are so hard (i.e., difficult) it's not easy to either create or explain the mathematical models to those in the field and hence the fields stagnate in the verbal or prescientific state, which allows everyone to have a 'theory' since any number of words strung together now comprises a theory. Even the discussion of what is a theory is constrained to discussion via classical bivalent logic or at best of statistical methods. One often encounters the terms equilibrium, stability and steady-state in connection with differential equations. In words, stable systems are those that return to their rest positions when the forcing is removed which implies that the response of the system to a disturbance is to return to its original position after some transients. This implies that there are no more changes in the system after a sufficient amount of time has elapsed. For the homogeneous case, the solution goes to zero as the independent variable (usually time) increases without bound. In general an nth order differential equation with constant coefficients can be written as a first order vector differential equation. In this case, a system is stable if the real parts of the eigenvalues of the matrix of coefficients are all negative. In some systems, the solution does not decay to zero; neither does it increase or decrease. In addition the concept of stationarity is usually assumed in connection with many stochastic processes. Stationarity implies that the statistics associated with the process do not change in time. Obviously, this means that the probability density is not an explicit function of time. The language evolution problem certainly doesn't fit into this mold, since the density that is derivable from the set of differential equations describing the language's evolution will be a function of time in general. However, there might exist a steady-state or a stationary version of the density. One final concept, that is often implicitly (and maybe unconsciously) used is that of ergodicity. Simply put, it means that the space (ensemble) averages of a stochastic process will be equal to its time averages. This is an additional concept piggybacked onto stationarity for making problems tractable. One of the ways in which random processes can be modeled, especially if there's a deterministic component which can be expressed as a differential equation, is via the Fokker-Planck method. If model is given by a set of differential equations

\[ \frac{dx_i}{dt} = f_i(x, t) + G(x, t) \omega(t) \quad i = 1, 2, 3 \ldots \]  

(A8.1)

where \( x \) and \( f \) are vectors, \( G \) is an \( n \times n \) matrix and \( \omega(t) \) is an \( m \)-vector zero-mean white Gaussian noise with the autocorrelation function;

\[ \langle \omega(t)\omega(t-z) \rangle = Q(t) \delta(t-z) \]  

(A8.2)

where \( Q(t) \) is positive semidefinite matrix and \( \delta(t) \) is the Dirac delta function, one can obtain an equatin for the first-order and second-order probability density functions as below;

\[ \frac{\partial \rho}{\partial t} = -\sum \frac{\partial}{\partial x_i} (\rho f_i) + \frac{1}{2} \sum \frac{\partial^2}{\partial x_i \partial x_j} \rho \ln \rho \]  

(A8.3)

For the special case of a single equation the above reduces to

\[ \frac{\partial \rho}{\partial t} = \frac{\partial}{\partial x} \rho(x, t) f(x, t) + \frac{1}{2} \frac{\partial^2}{\partial x^2} \ln \rho(x, t) \]  

(A8.4)

The Equations (A8.3) and (A8.4) are known as the Fokker-Planck equations or the forward Kolmogorov equations for the random differential equations shown. A general ordinary set of differential equations may be written as

\[ \dot{u} = F(u_1, u_2, \ldots, u_n, t) \]  

(A8.5)
where $= 1,2,...,n$ and prime indicates derivative. The $u$ may be thought of as the various intensive or extensive parameters describing evolution of a process. In general, the differential equation is a local description of the phenomenon since the effects of the various properties of the system on each other is described in an instant of time in a localized small region of the phase space. In contrast, the solution gives a global description of the behavior of the system. It is for this reason, that we seek solutions to differential equations. Finally, an important reason for this kind of a stochastic model is that it has both deterministic and random components. Even with the simplifications introduced, there are still useful analogies. For example, Gell-Mann, in the discussion of language as a complex adaptive system makes the statement (Hawkins & Gell-Mann, 1989)

In every one of these complex adaptive systems, there are fundamental rules, frozen accidents and selection pressures for functional outcomes. As usual, what is effectively a fundamental rule on one time scale may be a result of accident or functional adaptation in one of the complex adaptive systems operating on a longer time scale.

We can explain these concepts within the framework of the differential equation model. The Initial Conditions (IC) are one example of the frozen accidents or founder effects. Nonlinear equations exhibit behavior such as catastrophic (fast, discontinuous) changes and bifurcations. Nonlinear equations are also highly sensitive to initial conditions which is a hallmark of chaotic behavior. Since reality is always more complex than our simplifications, it often times behooves us to treat the simplistic deterministic systems as stochastic systems. The stochastic differential equation models are particularly useful since they show a deterministic component as well as a random one piggybacked onto the deterministic component. According to Siski, there are three ways in which randomness can enter into differential equations (Siski-1967, Van Kampen-1976 or Hubey, 1993).

(1) Random Initial or Boundary Conditions
(2) Random Forcing
(3) Random Coefficients

The randomness in the initial conditions is the easiest case since the mapping from the initial conditions to anytime which is accomplished by the Green's function is deterministic and if the source term is also deterministic, this implies simply that the probability density of the solution is a deterministic function of a random variable (the IC) and we can obtain the probability density of the solution (the IC part) via well known methods as can be seen, for example, in Papoulis (1984). The randomness in the source term means that there's a deterministic mapping (i.e. the Green's function) of a random function (the random part of the source or the forcing) to produce the final result. We can obtain some statistics about this process in many different ways, and has been extensively discussed in the literature.

The third case however is much more complex. A randomness in the coefficients implies that the Green's function itself is random and that there is now a random mapping (Green's function) of a random function (source). Thus the output $y(t)$ is comprised of at least two, possibly independent random processes. In the first case (i.e. randomness in the initial conditions) since the mapping from the set of initial conditions to the solution is deterministic, it essentially involves the calculation of the probability density of the solution from the probability density of the initial conditions using well known transformations. In the second case, the problem revolves around finding the deterministic solution of the equation. If the equation is linear, the steady-state solution can be expressed as an integral with the Green's function as the kernel of the integral. Thus the solution of the random problem involves a deterministic mapping of the random source and various methods such as mean-square analysis and statistics such as the autocorrelation and the spectral density of the process can be computed. If the equation is
nonlinear, then either linearization or Fokker-Planck methods may be attempted. The random coefficients problem is the most intractable. Methods for solving problems of this type can be found in Van Kampen (1978) among others.

Man came first to the realm of minerals, and from them he fell in among plants... when he left the plants and joined the animals, he remembered nothing of the vegetative state... In the same way he passed from realm to realm, until now he's intelligent, knowledgeable, and strong. He remembers not his first intellects, and he will leave this present intellect behind. Rumi (Chittick, 1983)

As a concrete example of the determinism in evolution, let's look at a model of random genetic drift in the narrow sense due to Kimura (Kojima, 1970). For equations of this type see also Roughgarden (1979). We assume that mutation, migration, and selection are absent, and that the change of gene frequencies from generation to generation is caused only by the random sampling of gametes in reproduction. We consider a locus in which a pair of alleles $A_1$ and $A_2$ are segregating. Using the Fokker-Planck method the forward Kolmogorov equation for the transition probability density $\phi(x,t)$ that the frequency (relative proportion in the population) of $A_1$ lies in between $x$ and $x + dx$ at time $t$, given that it is $\phi$ at the start (i.e. at time $t = 0$) is given by

$$\frac{d\phi}{dt} = \frac{3}{3x} f(x,t) \phi(x,t) + \frac{1}{2} \frac{d^2}{dx^2} [ g^2(x,t) \phi(x,t) ] = -\frac{3}{3x} M_{\delta x} \phi(x,t) + \frac{1}{2} \frac{d^2}{dx^2} [ V_{\delta x} \phi(x,t) ] \quad (A8.6)$$

where $V_{\delta x}$ and $M_{\delta x}$ are respectively the mean and variance of $\delta x$, the amount of change in gene frequency $x$ per generation. Now for the specific case in which a pair of alleles $A_1$ and $A_2$ segregating with respective frequencies $x$ and $1 - x$ in a random mating population of $N$ monogamous individuals and the mode of reproduction is such that $N$ male and $N$ female gametes are drawn as random samples from the population to form the next generation. Then the mean and variance in the change of gene frequency are

$$M_{\delta x} = 0 \quad \quad (A8.7a)$$

$$V_{\delta x} = x(1 - x)/(2N) \quad \quad (A8.7b)$$

Then the forward equation becomes (Kojima, 1970, 183)

$$\frac{d\phi}{dt} = \frac{1}{4N} \frac{d^2}{dx^2} (x(1 - x) \phi(x,t)) \quad \quad (A8.8)$$

Now, it's interesting that this equation is the same equation as the one for the probability density of the stochastic differential equation

$$\frac{dx}{dt} = g(x,t) \nu(t) = \left( \frac{1}{4N} \right) x(1 - x) \nu(t) \quad \quad (A8.9)$$

where $\nu(t)$ is a zero-mean white Gaussian noise. It's also interesting that Equation (A8.9) without the noise (i.e. $\nu(t) = 1$) is the famous logistic equation the discrete version of which has now become known as the Feigenbaum Oscillator of chaos fame. What this means is that we have a deterministic process whose coefficients are being multiplied by a random function. The coefficient $(1/4N)$ is essentially the propensity to segregate in one direction or the other. For a more complicated case in which we consider a reversible mutation rate $\nu$
from $A_1$ to $A_2$ and $v$ in the reverse direction, with $x$ the frequency of $A_1$, we have (Kojima, 1970:186)

$$M_x = ux + v(1-x)$$  \hspace{1cm} (A8.10a)

$$V_x = x(1-x)\left(\frac{1}{N_x}\right)$$  \hspace{1cm} (A8.10b)

From Equation (A8.3) we can see that they describe the stochastic differential equation

$$\frac{dx}{dt} = v(1-x) - ux + \left(\frac{1}{2N_x}\right)x(1-x)\omega(t)$$  \hspace{1cm} (A8.11)

Looking at this equation, it's difficult to use the words deterministic and random in the usual fuzzy sense in which it is used in biology, evolution and psychology. Yes the equation above has both deterministic and random components. And it cannot really be otherwise; no human can give birth to chimps or dogs. It's as deterministic as $2+2=4$ from this perspective. However we should note that the coefficients of the equations are considered to be random even if the equation itself without the added noise is a deterministic equation. In this case, there are other methods that can be used to derive solutions in different ways. There is an important result from Van Kampen (1976) that the average of a solution is not the same as the solution of averages. In other words if we have the equations for the evolution of a group of genes which determine a specific trait (i.e. polygenic trait) then we cannot simply solve the equation and find the average of the solution. This has importance in scaling from the micro viewpoint to the macro viewpoint when the equations are nonlinear and we have more knowledge of the micro behavior (i.e. specific genes) instead of evolution of the whole organism (i.e. macroscopic or polygenic traits). Therefore in order to produce equations for the evolution of a polygenic trait, we need to scale and average upwards from the equations [if we have them] for a set of genes. Similarly we'd have to scale and average upwards again from several polygenic traits toward higher level traits so on until we have some kind of averaged equations determining evolution of the whole organism.

In short, life implies intelligence. The absolute intelligence scale must come from living beings. Intelligence, order, organization, entropy, and probability are related concepts. Hence evolution is probably the best way to define intelligence, and evolution can be thought of as entropy which completes the circle. In truth, the link between information theory and probability theory (and between error and chaos) is probably the best link to enable one to model something as complex as society. In any case, the intelligences of animals (including humans, of course) do seem to fulfill the conditions for scalability of Guttman (see for example, Maranell, 1974). As for an absolute or a ratio scale for intelligence, it remains yet to be seen what can be used except perhaps some kind of a brain-size to body-size ratio.

A plot of the DNA nucleotide pairs as a percent of the mammalian DNA content shows a remarkable correlation with what we consider using common sense to be levels of intelligence. (Britten and Davidson, 1969). A log-log plot of the time of origin of the species (years ago) against the number of bits of information encoded in the DNA nucleotide pairs (per haploid cell) (borrowed from Britten and Davidson), and also against the information content of the brain, in Sagan (1977) also shows a correlation with intelligence. Similarly a plot of brain mass against body mass of mammals (and also of all animals) shows an expected ordering according to intelligence (Sagan, 1977). One can see similar patterns or order in the brain masses and also of the parts of the brain associated with higher functions in the primates in Eccles (1989). The overall conclusion is that an absolute scale of intelligence should in some way be associated with comparison of brain mass to body mass. It should probably be of the form
\[ I = W \left( \frac{b - kB - c}{B} \right) = W \left( \frac{b - c}{B} - k \right) = W \left( \frac{d}{B} - k \right) \]  

(A8.12)  

where \( b \) = brain mass, \( B \) = body mass, and \( k \) and \( c \) are constants. The reason for subtracting a constant amount from \( b \) is that there must be some minimal amount of brain matter necessary to keep the life processes going. The reason for subtracting an amount proportional to the body mass \( B \) is that there is probably a need as the body size gets larger for greater or finer control of the voluntary muscles so that an elephant or a whale would seem to need more neurons purely for its existence and not for higher or more abstract levels of cognition. The form of the function \( W() \) itself is of great concern but it would seem to be almost arbitrary in the sense that, we'd have to decide with our naked senses that, say we are more intelligent than chimps, and that chimps are more intelligent than dogs, and so on. We might note, for example, that chimps seem to recognize themselves in mirrors. As another example, we might note that a dog or a cat would notice and try to shake off something stuck to its paws whereas a duck might be oblivious to it. We might also indirectly try testing the number sense of various animals.

There really is not that much difference in ordering the intelligences of animals according to some minimal types of behavior patterns than noticing that as the temperature gets hotter the liquid in a glass bulb thermometer goes up and not down, or that the things that seem to be heavier stretch a spring more than those that aren't. We have further hints that the proportion of masses as an indication of intelligence (at least across species) is in the right direction by noting that dimensional analysis which is often used to obtain the forms of the parameters in dealing with complex areas such as fluid dynamics dictates that we use a ratio of the masses so as to make the term dimensionless. The usefulness of dimensional analysis can also be seen in the construction of metric spaces for speech analysis in Hubey (1994).

We could, of course, try a logarithmic scale something like the decibel scale using a ratio of human intelligence (or the brain/body ratio as above) to the intelligence of the animal in question. Whatever the case, it is important that an absolute/ratio intelligence scale be used that can pass muster of researchers in artificial intelligence, psychologists, and neurophysicists. And if intelligence is to be broken down into its components (however many) then there should be an operational definition of a single number that is derived from such tests. In such a case we can use a weighted function of the various parts of the brain so that some kind of an absolute intelligence scale should be established with which we can make progress toward machine intelligence instead of simply being forced to deal with unrelated and incoherent ideas such IQ tests using correlation-regression analysis, Turing tests for machine intelligence, expert systems whose knowledge (and intelligence) defy our imagination.