Can Intelligence be Measured?

Geoffrey Marnell

It is perhaps a platitude to say that intelligence is highly regarded in most societies. The wooden-headed are invariably scorned and often mocked. They are shepherded into special classes at school and, later in life they, find themselves in dull jobs. Their opinions are mostly spurned—other than as the butt of jokes—and they tend to remain on the periphery of power and prestige. Better by far to be—or be seen as—intelligent. So it is hardly surprising to read that:

> "scores on intelligence related tests matter, and the stakes can be high."[1]

Intelligence is one of society's most prominent discriminators. Age, fitness and gender are perhaps the only discriminators more prominent, at least in secular countries. There is some justification for age discrimination (at least at the lower end of the scale). However much a ten-year-old wants to work in a mine rather than study at school, there are good reasons why society is justified in forbidding that. The discrimination against an amputee wanting to work as an airline pilot also appears justified, though obviously for different reasons. Discrimination based on gender is not justified under any circumstances (a fact not annulled by our seeming inability to eradicate it).

What of intelligence? It would seem unfair to deny someone a job where intellectual prowess is not needed solely on the grounds that their score on some test was to the left of the bell curve of general intelligence. But just as some jobs require limbs, others require a sharpness of mind not possessed by everyone. A neurosurgeon who encounters an unexpected bleed while debulking a tumour might need to make a split-second decision if the patient is to be saved, as might a pilot encountering an unexpected intrusion into allocated airspace. A society as complex, and expectant, as ours needs quick-thinking highly intelligent members. Our welfare depends on it.

Society tends to reward the more intelligent more than the less intelligent (although that is not always the case: a dim-witted footballer can earn millions a year while a sharp-minded scientist might earn a tenth as much). The reasons are not always obvious or straightforward. But whatever they are, rewards are finite and the demand for them is invariably greater than the size of the pot. So some form of discrimination is necessary. A country of 20 million people cannot afford to have a million neurosurgeons or airline pilots. The discriminant mostly used to select who gets rewarded and who misses out is a blend of knowledge and intelligence (although some professions will call for other attributes as well). And in that blend, intelligence is usually given precedence. The ten trainee pilots applying for the one job might have had exactly the same training—and thus the same knowledge—but only one can be chosen. One hundred students might have been taught exactly the same things and thus have exactly the same knowledge, but if they are all applying for a single scholarship, then obviously something more than just knowledge is needed as the discriminant. In such cases, the discriminant society has chosen is usually intelligence, and our degree of intelligence is usually computed from the results of an intelligence test we take.

> "testing of abilities has always been intended as an impartial way … of determining who gets what [such as] power and privilege, responsibility and reward."[2]

> "In the United States today, high test scores and grades are prerequisites for entry into many careers and professions."[3]

1. U. Neisser et. al., "Intelligence: Knowns and Unknowns", *American Psychologist*, vol. 51, No. 2, 1996, p. 78.
2. L. J. Cronbach, *Essentials of Psychological Testing*, Harper & bRow, New York, 1984, p. 5.

Since our future could be determined in some not inconsiderable way by the score we are given on an intelligence test, it is critical that intelligence tests really do measure the thing that society expects them to measure: intelligence. If someone is to be denied power, privilege, responsibility or reward on the basis of a test, the test had better be a good one.

But what is intelligence? Alas, it seems that not even cognitive psychologists agree on what it is:

> "when two dozen prominent theorists were recently asked to define intelligence, they gave two dozen somewhat different definitions"[4]

If there is no strong agreement on what intelligence is, how can we be confident that any particular intelligence test is actually measuring intelligence? Perhaps the tests are good at predicting achievement at school—although that correlation is only 0.5[5]—but does that warrant them being called tests of intelligence (with all the highly charged connotation that that word carries)?

That doubt can be magnified by a close examination of some of the questions asked on standard intelligence tests. It seems in keeping with the common-or-garden idea of intelligence that the number of correct answers given on an intelligence test would be the prime measure of intelligence: the more correct answers, the more intelligent. That is no doubt the assumption of those who are being tested. But it is not the case, as I'm about to show.

Psychologists will claim that it is "a weakness [if a question on an intelligence test allows] the putting of an ingenious alternative answer by some high-grade subjects".[6] Yet many intelligence tests pose questions for which there is not one correct answer, other than the answer that only a "high-grade subject" would give. The type of question I have in mind is the number-sequence problem, a very simple example of which is {2, 4, 6, 8, 10, ?}. One answer, and the one that would no doubt be marked correct, is 12. An answer of 13 would probably be interpreted as a hurried slip of the pen; an answer of –3243 is likely to be interpreted as evidence of cognitive perversity. Yet all three answers (12, 13 and –3243) are equally correct, correct in the sense that there is a knowable rule that will generate the sequence as continued. In fact, as I am about to show, any number can continue the above sequence—can continue any finite sequence—and the sequence so continued will be the product of a strict mathematical rule (an equation, in other words—a polynomial, to be precise). So there will be an infinite number of such rules capable of generating any given sequence. In other words, there is an infinite number of solutions to any number-sequence problem. So, a "high-grade subject" knowing this to be the case, might give the correct answer— namely, "any number"—and be marked wrong. Their answer will be marked wrong if they are responding to a multiple-choice question (where the assumption is explicit that there is only one correct answer and it is a number). It will be marked wrong if the marking is done not by a human but by an optical scanner. And it will be marked wrong if marked by a human who is merely following the answers provided by the test designers. In other words, the correct answer is incorrect. Is that intelligent?

Psychologists have long considered number-sequence problems to be particularly useful in intelligence testing, some even suggesting that such problems "provide one of the best tests of reasoning".[7] Not all intelligence tests include such problems, but many do.[8] But if they admit to an infinite number of answers—equivalent to one correct answer: "any number"—how useful can they be?

*

3. U. Neisser op. cit., p. 96.
4. ibid., p. 77.
5. ibid., p. 96.

Let me give a short inductive proof that a finite sequence can be continued in any way and the sequence as continued can be generated by a simple mathematical rule (a polynomial). The proof—which requires no sophisticated mathematical knowledge—follows from a curious feature of all polynomials. (If you are not keen to revisit a little secondary school maths, skip to the next section.)

A polynomial is an algebraic expression of the general form:

$$y = ax^n + bx^{n-1} + cx^{n-2} + \ldots + gx + h$$

One sort of polynomial we will all be familiar with—probably through years of dreary classroom factorisation—is the quadratic equation (for example, $y = x^2 + 2x + 3$). A quadratic equation is said to be a polynomial of degree 2, since 2 is the highest power in such an equation. Polynomials range from degree zero (for example, $y = 3$) to any positive degree you might wish to imagine. For our discussion it is important to note that a number sequence can be generated by repeatedly plugging consecutive x values into a polynomial and solving for y. Using the quadratic mentioned above, plugging in one value at a time from the set of consecutive x values {0, 1, 2, 3, 4, 5 …} yields the number sequence {3, 6, 11, 18, 27, 38 …}.

Now to an interesting feature of all polynomials: a sequence of numbers derived from any polynomial is reducible, by the repeated subtraction of adjacent pairs, to a zero sequence. Let me explain this using, as an example, the quadratic polynomial mentioned above, namely $y = x^2 + 2x + 3$. As we noted, the sequence of y values {3, 6, 11, 18, 27, 38 ...} can be derived by substituting consecutive integers for x in this equation starting with $x = 0$. Taking each pair of numbers in turn and subtracting the number on the left from the number on the right yields a further sequence: {3, 5, 7, 9, 11 ...}. Repeating this procedure but this time using the figures in the newly generated sequence gives the sequence {2, 2, 2, 2 ...}. Repeating the procedure again yields a zero sequence: {0, 0, 0 ...}. This process of repeated subtraction can be displayed graphically like this:

y:	**3**		**6**		**11**		**18**		**27**		**38**
(1)		3		5		7		9		11	
(2)			2		2		2		2		
(3)				0		0		0			
(4)					0		0				
(5)						0					

where each number in a subtraction sequence—any of those labelled (1) to (5) at the left—is derived by the subtraction of the two figures immediately above it. I will call this lay-out a *subtraction triangle*. Obviously a subtraction triangle can be generated for any given sequence of numbers. Further, it will run out of pairs after (t – 1) subtraction sequences, where t is the number of figures given in the original sequence. (Note that 6 numbers are given in the original sequence above and there are no more pairs after the fifth subtraction sequence.) I will call the last figure in any particular subtraction triangle—the one without a pair—the *concluding term*.

Any polynomial, whatever its degree, generates a sequence that is reducible in this way to a zero sequence—that is, a sequence solely of zeros—and the number of steps (that is, subtraction sequences) needed to reach a zero sequence is equal to (n + 1), where n is the degree of the polynomial. (A quadratic polynomial has degree 2 and so 3 steps are required to reach a zero sequence from any sequence generated by a quadratic equation, as shown by the (3) in the above subtraction triangle beside the first zero sequence. A cubic equation is of degree 3 and would require 4 steps, and so on.)

6. E. Anstey, *Psychological Tests*, Thomas Nelson, London, 1966, p. 162
7. P. E. Vernon, *Intelligence and Attainment Tests*, University of London Press, London, 1960, p. 81.

Just as we can derive the subtraction triangle for a sequence generated by a specific polynomial—as we just did—we can derive the subtraction triangle for the general sequence derivable from the general form of some polynomial. For example, the general form of polynomials of degree 2—quadratic equations, in other words—is $y = ax^2 + bx + c$. From this the following subtraction triangle can be derived. (For reasons of space, I have split this general subtraction triangle into two parts, representing x values of {0, 1, 2, 3, 4 ...} and {... (n – 3), (n – 2), (n – 1), n } respectively.)

Part (a)

x:	**0**		**1**		**2**		**3**		**4**	**...**
y:	**c**		**a + b + c**		**4a + 2b + c**		**9a + 3b + c**		**16a + 4b + c**	
(1)		a + b		3a + b		5a + b		7a + b		
(2)			2a		2a		2a		[2a]	
(3)				0		0		0		

Part (b)

x:	**n – 3**		**n – 2**		**n – 1**		**n**
y:	**$a(n–3)^2+b(n–3)+c$**		**$a(n–2)^2+b(n–2)+c$**		**$a(n–1)^2+b(n–1)+c$**		**an^2+bn+c**
(1)		2an–5a+b		2an–3a+b		2an–a+b	
(2)	[2a]		2a		2a		2a
(3)		0		0		0	

Figure 1. Subtraction triangle for the general form of the quadratic equation

This shows that a sequence generated by a quadratic—any quadratic—yields, at the third level of subtraction, a zero sequence, and does so irrespective of how far one extends the sequence (that is, irrespective of the value of n).

Note now that we can use this general subtraction triangle as a template to easily derive the equation that generates any sequence of y values we suspect can be fit to a quadratic equation. (This will be so if the first zero sequence occurs at the third subtraction level). Our template for polynomials of degree 2 shows that the first term in the given sequence is c. This is the third co-efficient in whatever equation generates the sequence. The first term in subtraction sequence (1) can be found by subtracting the first term in the y sequence from the second term. From our template we see that the first term in subtraction sequence (1) is equal to $a + b$. Continue subtracting adjacent terms in the y sequence to derive the second term in sequence (1). Knowing the first two terms in sequence (1) enables us to derive, by subtraction, the first term in sequence (2) which, by our template, is equal to $2a$. We can use this figure to derive a and then b (since the first term in sequence (1) is $a + b$). We now have all the co-efficients of a quadratic equation that can generate the given y sequence.

For example, {3, 11, 25, 45, 71 ...} are consecutive y values that can be generated by a quadratic equation. (You can tell this by deriving a zero sequence at subtraction level 3.) It follows from our quadratic template that $c = 3$, sequence (1) begins {8, ...}, $a + b = 8$, sequence (2) begins {6 ...} and hence $2a = 6$. So a = 3 making b equal to 5. Thus the generating quadratic is:

$$y = 3x^2 + 5x + 3$$

General subtraction triangles can be derived in like fashion for polynomials of any degree. The following table gives, for those of degree 1 to 6, the starting terms for each subtraction sequence as far as the first zero sequence. I have derived these equations by creating a

8. Some examples: the AQ, MQ and PQ tests developed by the Australian Council for Educational Research. In fact, 14 of the 34 questions on the PQ test are of the standard number-sequence variety, with another 9 being a slight variation on the theme.

subtraction triangle for the general form of each polynomial shown in the top row of the table (just as we did above for quadratics).

y =	ax + b	ax²+bx+c	ax³+bx²+cx+d	ax⁴+bx³+cx²+dx+e	ax⁵+bx⁴+cx³+dx²+ex+f	ax⁶+bx⁵+cx⁴+dx³+ex²+fx+g
Y:	b	c	d	e	f	g
(1)	a	a+b	a+b+c	a+b+c+d	a+b+c+d+e	a+b+c+d+e+f
(2)	0	2a	6a+2b	14a+6b+2c	30a+14b+6c+2d	62a+30b+14c+6d+2e
(3)		0	6a	36a+6b	150a+36b+6c	540a+150b+36c+6d
(4)			0	24a	240a+24b	1560a+240b+24c
(5)				0	120a	1800a+120b
(6)					0	720a
(7)						0

Figure 2. Template of starting terms for subtraction triangles of polynomials of degrees 1–6

This table of templates can help us quickly derive a polynomial responsible for generating any sequence of numbers of the sort found on typical intelligence tests. But let's put it to use in deriving the polynomial behind a seemingly random sequence. Take {–6, 0, 0, 0, 6, 24 ...} for instance. You might think that no mathematical rule could possibly produce such a mess of numbers. But you would be wrong. The associated subtraction triangle for this sequence reveals a zero sequence at subtraction level 4, with the subtraction sequences beginning 6, –6, 6 and 0 respectively. Hence a polynomial of degree 3 can generate this sequence. The equations for the starting terms of the subtraction sequences derived from polynomials of degree 3—given in the fourth column of figure 2 above—allow us to conclude that $6a = 6$, $6a + 2b = -6$, $a + b + c = 6$ and d = the first term in the given sequence. We now have a set of simple simultaneous equations to solve which shows us that the generating polynomial is $y = x^3 - 6x^2 + 11x - 6$. So the sequence is not so unruly after all.

Now to a proof that an infinite number of polynomials can generate any given sequence. To show this I'll derive the subtraction triangle corresponding to the sequence {1, 4, 9, 16 ...), although any sequence will give the same conclusion.

Y:	1		4		9		16	P
(1)		3		5		7		
(2)			2		2			
(3)				0		Q		

The first thing to note in this triangle is that there is a subtraction sequence that could be continued as a zero sequence, namely, sequence (3). Hence we are entitled to assume that there is a polynomial of degree 2 that could generate the given y sequence. Indeed, using our template for polynomials of degree 2—see figure 2 above—we can quickly derive the form of this polynomial, namely, $y = x^2 + 2x + 1$. We can easily check that this equation does indeed yield the given y sequence by plugging in consecutive values for x starting from zero and solving for y. We could also use this equation to calculate a possible next term in the sequence: 25.

It is important to understand why we are entitled to assume that subtraction sequence (3) is a zero sequence. We can do so because the only constraint we are under in creating the subtraction sequences—which, being sequences, can continue indefinitely—is to ensure that together they generate the numbers in the initial y sequence and in the given order. And the numbers in the sequence {1, 4, 9, 16 ...}, and their order, is in no way affected by our assuming that subtraction sequence (3) continues {0, 0, 0}. In other words, we are entitled to extend and change a subtraction triangle in any we please so long as the original y sequence remains unchanged.

Now just as we are entitled to assume that subtraction sequence (3) in the above subtraction triangle is a zero sequence, we are entitled to assume that it is not. Given the numbers we have in the y sequence, Q—the next term in sequence (3)—can be any value at all.

In other words, whatever value we give to Q, the part-sequence given—{1, 4, 9, 16 ...}—will not change.

Bearing this in mind, let's shift our attention from Q to P, the next term in the given y sequence. (In the typical number-sequence problem on an intelligence test, P is the number being asked for.) We have already found a polynomial to generate a value for P—namely, 25—but let's give P some other value. Any value, in fact: positive or negative. What we now find is that by continuing the subtraction triangle—creating lower-level subtraction sequences, in other words—another subtraction sequence can be found—or created—that can be assumed to be the beginning of a zero sequence. And this means that there must be another polynomial that can create the particular y sequence we are investigating. Let's see how this works by substituting, say, 13 for P in the y sequence above and continuing the subtraction triangle:

Y:	**1**			**4**			**9**			**16**			**13**				
(1)		3			5			7			−3						
(2)				2			2			−10							
(3)					0			−12									
(4)							−12			**R**							
(5)								**S**									

Our extended subtraction triangle runs out of puff at subtraction level (4), there being no partner for the new concluding term (–12) to be subtracted from. But there is no reason why we cannot continue subtraction sequence (4) and do so as we please, since doing so will not affect the given y sequence {1, 4, 9, 16 ...}. So let's make R equal to the current concluding term: –12. This enables us to create subtraction sequence (5), with S, our new concluding term, equal to zero. Because S is equal to zero we are entitled to assume that subtraction sequence (5) is a zero sequence. Hence there must also be a polynomial of degree 4 (that is, the number of the subtraction level corresponding to a zero sequence less 1) capable of generating a sequence starting {1, 4, 9, 16 ...}. This is in addition to the quadratic we uncovered earlier when we stopped at subtraction level 3. We can now apply the template for polynomials of degree 4 (shown earlier in figure 2) to derive the equation for this polynomial, which turns out to be:

$$y = -\frac{1}{2}x^4 + 3x^3 - \frac{9}{2}x^2 + 5x + 1$$

You can check that this equation does generate the sequence {1, 4, 9, 16, 13 ...} by plugging in consecutive x values beginning from zero and solving for y.

Of course, we didn't have to make $P = 13$. We picked a number just to show that there will, or can, be a zero sequence in the subtraction triangle, and a zero sequence always indicates that a polynomial can generate the top-level sequence. But the number we do make P will partly determine the value that starts subtraction sequence (4), and this term will partly determine the co-efficients of the corresponding polynomial. So there will be a unique polynomial for each possible value of P. Since P can be any number at all and a zero sequence found, there are any number of polynomials of degree 4 capable of generating the sequence {1, 4, 9, 16, ?}.

But the infinity of possibilities so far uncovered is but a drop in the ocean. For each additional term with which one continues the given y sequence there will be yet another infinity of polynomials capable of generating that sequence. For the term that follows P in the above sequence enables the subtraction triangle to be extended by one additional subtraction level. In other words, where P caused the concluding term to fall at the start of subtraction sequence (4), the term after P will cause the concluding term to fall at the start of subtraction sequence (5). Again we can continue the final subtraction sequence with identical terms, thus enabling us to create a further subtraction level—subtraction sequence (6)—which can be a zero sequence. This means that there must be polynomials of degree 5 capable of generating the y sequence as given—an infinity of them, in fact: one for each possible value that the term after P can take. And then, of course, there is the term after the term after P—generating an infinity of polynomials of degree 6. And so on ad infinitum.

Hence an infinite number of polynomials can generate any given number sequence.

The above technique can be used to uncover generating polynomials for any sequence, no matter how unruly or random the terms in the sequence appear, and irrespective, too, of how many numbers are given. Let's illustrate this using an unlikely sequence of numbers, a sequence I'll just pull out of the air: {3, –112, 69, 1003, –47 ...}. The corresponding subtraction triangle looks like this:

Y:	**3**		**–112**		**69**		**1003**		**–47**
(1)		–115		181		934		–1050	
(2)			296		753		–1984		
(3)				457		–2737			
(4)					–3194		**X**		
(5)						**Y**			

Without it affecting the given sequence, we can substitute –3194 for X and get a value of zero for Y. Hence can treat sequence (5) as a zero sequence. Since we have the beginnings of what could be a zero sequence, we know that some polynomial can generate the y sequence given. We also know that it is a polynomial of degree 4—since subtraction level (5) is a possible zero sequence—and so we know the appropriate template to apply to derive the particular equation. Applying that template gives us the following polynomial:

$$y = -\frac{1597}{12}x^4 + \frac{2624}{3}x^3 - \frac{18533}{12}x^2 + \frac{4127}{6}x + 3$$

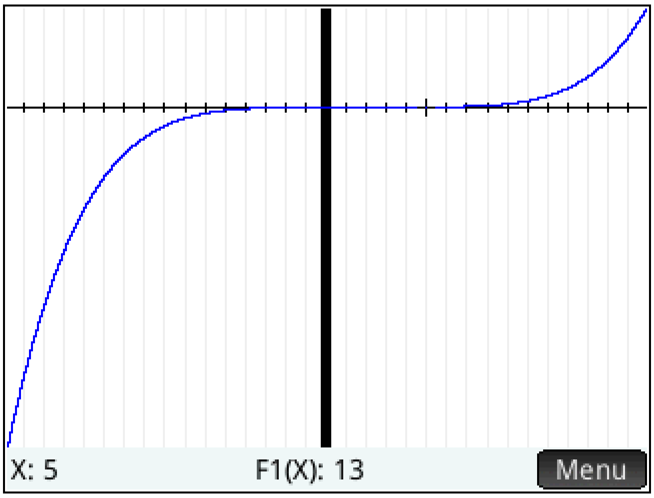
Plugging in consecutive values for x from the set {0, 1, 2, 3, 4 ...} will verify that this polynomial does indeed generate the sequence {3, –112, 69, 1003, –47 ...}. If we had made X in the substration triangle some value other than –3194, we could have continued subtraction sequence (5) by repeating whatever value Y had become. Then we could create a zero sequence as subtraction sequence (6) and derive a polynomial of degree 5. In fact, we could derive a unique polynomial of degree 5 for each possible X value.

Hence any given sequence is derivable from an infinity of polynomials, from which it follows that there is never just one solution to the problem of how a given sequence continues in an orderly way.

Returning to the sequence that opened this paper: if we continue it {2, 4, 6, 8, 10, 13 ...} and derive the associated subtraction triangle, we find a concluding term of 1 at subtraction level (5). Hence a zero sequence can be created at subtraction level (6), indicating the presence of a generating polynomial of degree 5. Our template for such polynomials gives us its exact form:

$$y = \frac{1}{120}x^4 - \frac{1}{12}x^4 + \frac{7}{24}x^3 - \frac{5}{12}x^2 + \frac{11}{5}x + 2$$

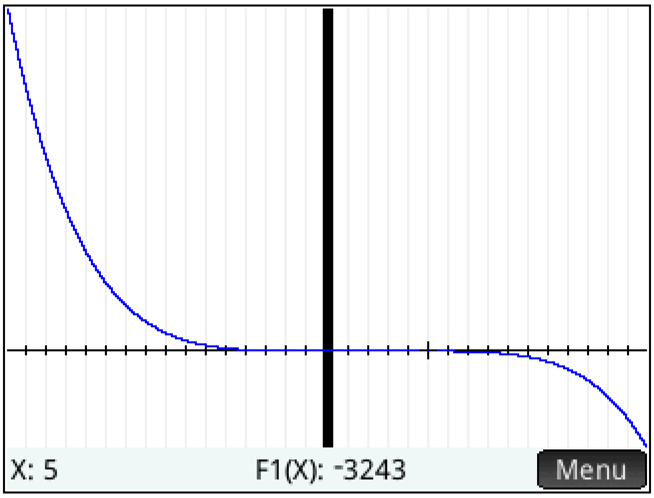
The sequence generated by this polynomial is {2, 4, 6, 8, 10, 13, 20, 37, 74, 146 ...}, and if we graphed it, it would look entirely harmless:



If we had continued the sequence with 14, not 13, we would find yet another polynomial of degree 5 capable of generating the given sequence, and so on. And what about that seemingly perverse answer mentioned earlier, namely {2, 4, 6, 8, 10, –3243, ...}? Well, here is the formula:

$$y = -\frac{217}{8}x^5 + \frac{1085}{4}x^4 - \frac{7595}{8}x^3 + \frac{5425}{4}x^2 - 649x + 2$$

And here is the corresponding graph (nowhere near as skittish and disorderly as you might have thought):

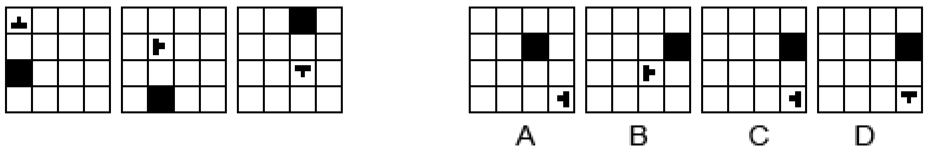


What reasons are there, then, for marking any answer to a number-sequence problem as wrong? If a problem admits to many answers and yet provides no objective way of ranking the possible answers in a way that assists the measuring of what the problem was designed to measure then surely it cannot be a valid test item. Consequently, such problems should be banished from intelligence tests.

Number-sequence problems appear to be testing both numerical and pattern-recognition abilities. But the problem of multiple answers can also bedevil questions that attempt to assess visual pattern-recognition skill. Consider the following question from an intelligence test[9]:

9. This question is taken from a sample intelligence test published by IQ Test Labs. See http://www.intelligencetest.com/index.htm. (Viewed 23 March 2014)

2. Find the picture that follows logically from the diagrams to the right.



The answer given is C. One way to arrive at C is if the following algorithm is applied to one frame to generate the next:

- each object—the black block and the perpendicularity symbol—is moved one column to the right
- each object is moved down one row
- if such movements would take an object below the grid it is placed in the top row, and if such movements would take an object to the right of the grid it is placed in the first column, and
- each object is rotated 90°clockwise.

For the sake of simplicity, let's call jumping to the top row *cycling*.

Now why is that algorithm more logical than the following?

- the perpendicularity symbol is moved one column to the right
- if no cycling occurred at the last move, the black block is moved one column to the right, otherwise it remains where it is
- each object is moved down one row
- if such movements would take an object below the grid it is placed on the top row, and if such movements would take an object to the right of the grid it is placed in the first column, and
- each object is rotated 90°clockwise.

This equally logical algorithm makes A the correct answer. And no doubt other algorithms—none of which are illogical, whatever that might mean—would make B or D the correct answer.

*

When the ambiguity of a question on an intelligence test is pointed out, three defences are often put forward:

- The answer being sought is the logical or most logical of all the possible answers.
- The answer being sought is the simplest or most obvious of all the possible answers.
- It doesn't matter what answer is given so long as it is the answer given by most people who have been independently verified as highly intelligent.

The first defence is just plain silly. It makes no sense to say that a polynomial can be illogical or is more logical than another polynomial. Likewise with algorithms that account for a series of visual patterns. If one computer programmer uses, say, the Miller–Rabin primality test to determine if a number is prime and another programmer uses the Baillie–PSW primality test, it would be bizarre to call one programmer's algorithm logical or more logical than the other. Perhaps one algorithm is more efficient than the other—it might take less processing time—but efficiency is not the same as logical. (Is a company that can create

100 widgets with 10 staff less logical than a company that can make 112 widgets with 9 staff?) Logical is an absolute adjective. It has no comparative or superlative forms.

The second defence is the position of the late Victor Serebriakoff (co-founder of Mensa and great fan of intelligence tests), set out in in a personal correspondence. That intelligence tests never explicitly ask for the simplest or most obvious answers may be in recognition of the fact that the meaning of *simplicity* and *obviousness* is neither simple nor obvious.

Consider, firstly, the meaning of *simple*. In the context, say, of answering an arithmetical question on an intelligence test—for example, {0, 1, 3, 6, 10, ?}—there seems to be only two likely candidates: (a) simple in the sense of requiring few steps or (b) simple in the sense of being easy.

By the first definition, one answer to a question will be simpler than another if it was reached with fewer steps. But suppose we have two possible answers to an arithmetical question, both employing the same number of steps but one using more-advanced arithmetic. An example is {2, 4, 16, 128, x}. One answer is 1,240, derived by noting, firstly, that the numbers increase by {2, 12, 112, …} and, secondly, that these differences increase by {10, 100, …}, a set that could conceivably continue with 1,000. So if $x - 128 = 1112$, then $x = 1{,}240$. Another possible answer is 2,048, derived by noting that the numbers in the sequence are all powers of 2, the indices form the set {1, 2, 4, 7, …} and that the differences between consecutive indices forms the sequence {1, 2, 3, …}, a sequence which could conceivably continue with 4. And 2^{11} is 2,048. Note that both answers require two steps: in the first case we derive one set of differences and then another; in the second we derive indices and then the set of differences between those indices.

But are the two answers as simple as one another? The answer is surely no, for powers are certainly more complex than subtraction (which is why powers are taught later than subtraction at school). Hence the first answer in our example (1,240) would have to be considered simpler than the second even though it needed just as many steps.

Hence treating simplicity as a measure of the number of steps involved in reaching an answer is inadequate. We need also to be able to rank the various types of arithmetical operations in terms of some other sort of simplicity, and the most likely candidate appears to be simplicity as easiness.

But now the waters cloud with a good deal of subjective pollution. Obviously powers should, on our new definition, be ranked as less simple than subtraction. But is multiplication less simple than addition? Perhaps it is simpler than division, but is it simpler than powers? (Aren't powers just a form of multiplication?) And doesn't it all depend on the size of the numbers being manipulated? For example, 23,456,876 plus 213,546,765 might not be as simple as 2 times 3, but it is simpler than 234,321 times 216,876.

It's now getting messy and complicated and, to make matters worse, the rankings derived will also need to be weighted. For if, say, addition is simpler than powers, might it nevertheless be the case that an answer requiring two steps of addition is less simple than an answer requiring just one step of powers? Similarly, might not two steps of addition each involving three-digit numbers be less complex than three steps of addition each involving two-digit numbers?

This suggests that the arithmetical simplicity of an answer is a compound of simplicity in both senses outlined earlier. In other words, it is a function both of the number of steps taken to reach the answer and the relative easiness of the operation used in each step. For the sake of an illustration, let's say that single-digit addition is given an easiness weighting of 1.0, two-digit addition a weighting of 1.2 and powers to base 2 a weighting of 2.3. It would follow from these weightings that an answer involving one step of powers to base 2 is less simple than an answer involving a step of single-digit addition and a step of two-digit addition.

But no mathematician has established—or is ever likely to establish—a table of objective rankings and weightings for all the arithmetical operations (taking into account the various degrees of complexity due to the various sizes numbers can take). The task would be daunting and the objectivity doubtful. But without this information, how can we ever prove that the x steps used in deriving answer p indicate an answer simpler than q derived using y steps?

Even if some objective way of quantifying this concept could be found, we would still need to radically rethink the way intelligence tests are administered and marked. For if simplicity is partly a function of the number of steps needed to derive an answer, a criterion that links correctness to the simplest answer means that testers need to see not just the answer to a question but also the method the testee used to derive it. For if two testees arrived at the same answer but one took more steps, then the answer derived using more steps would need to be marked as incorrect even though it is identical to the answer offered by the other testee. That testees are not asked for their methods as well as their answers suggests that psychologists do not consider simplicity to be an important factor in determining the acceptability of answers to questions on intelligence tests.

Consider, now, the concept of obviousness. The popular British puzzlists Gyles Brandreth and Trevor Truran proposed this devilish number-sequence problem in one of their many books: {1, 8, 3, 7, 1, 9, 0, ?}. If you had a passion for British history, an obvious answer might be 1, the very answer the authors give (1837–1901 being the reign of Queen Victoria). That this would hardly be considered obvious by many others indicates, perhaps, that obviousness is an intractably relative concept. A trivial example, perhaps, so let's consider an example where numerical reasoning is expected: {4, 8, 32, 512, ?}.Once again the obvious answer is not so obvious. To some, the numbers in this sequence will be seen immediately as powers to base 2, with the indices forming the sequence {2, 3, 5, 9 …}. The differences between these indices form the sequence {1, 2, 4 …}. Is it obvious, now, how this new sequence continues? Is each number twice the previous number, leading to 8 as the continuation? Or is the sequence proceeding by progressive addition (+1, +2, +3 …) leading to 7 as the continuation? It is certainly not obvious which of these continuations is the most obvious.

Of course, some would not see {4, 8, 32, 512 …} as a set of powers to base 2. They might see each number as half of the square of the previous number. Others might think that adjacent pairs are being multiplied together and the product multiplied by the corresponding number in the sequence {1, 2, 3, 4 …). Here, then, are four non-obscure ways of continuing {4, 8, 32, 512 …} each of which could reasonably be considered by those who adopted them as obvious. But how are we to judge which is the *most* obvious?

In one sense it makes no sense to talk of something being more obvious than something else. Something is either obvious or it is not. The word *obvious*—like the word *logical*—is an absolute adjective: it has no comparative or superlative forms. In this sense, something is obvious if it is apparent or evident. If someone who has given 3 as the square root of 9 is told that –3 is also a square root of 9, –3 becomes an obvious answer. But it is not less obvious than 3 for not having been known at the time the question was asked. An appropriate parallel is this: answers to the question "where can I ski in Australia?" might include "Thredbo" and "Falls Creek". To someone who knows that you can ski at these places, the two answers are obvious. But one answer is not more obvious than another: they are just both obvious. Suppose, now, that someone answered this question with "Perisher". To someone who knew you can ski at Perisher, this answer is obvious; to someone who didn't, it is not. But once they know you can ski at Perisher, "Perisher" becomes an obvious answer. Further, it wasn't a less obvious answer beforehand; it just wasn't obvious at all. To return to sequences: two people might each see {4, 8, 32, 512 …} as a set of powers to base 2 with the indices forming the sequence {2, 3, 5, 9 …} and the differences between the indices forming the sequence {1, 2, 4 …}. They might then interpret the latter sequence differently. One might give 128 as the answer (if the sequence of differences is interpreted as continuing with 7) and the other might answer 256 (if the sequence of differences is interpreted as continuing with 8). Each might think that their own answer is obvious. But when told the other's answer, each is more likely to say "That wasn't obvious!" rather than "That is less [or more] obvious than the answer I got!".

It might be retorted here that "Perisher" is a less obvious answer because it is not as widely known that you can ski at Perisher as it is that you can ski at, say, Thredbo. This introduces a different sense of obvious. In this sense, a more obvious answer is one that is more widely known, and the most obvious answer would be the one most people would give.

Applied to intelligence tests, this new sense of obvious suggests that test constructors should give a particular problem to a sample of people and declare the most obvious answer to be the one that most people in that sample arrived at. From then on that answer will be the accepted answer.

But this approach is highly questionable. It is axiomatic that the highly gifted are disposed to give answers to test questions that the majority do not give. (That is partly what sets them apart.) Indeed, such people will often deliberately give a more obscure answer to an intelligence test question thinking that more marks might be awarded for being bright enough to detect an answer more complex but equally correct. To penalise the gifted for doing so is contrary to one of the purposes of intelligence tests: to identify those of superior intellect.

The third defence given to allowing questions on intelligence tests to have more than one answer is that it doesn't matter what answer is given so long as it is the answer given by most people who have been independently verified as highly intelligent. This is the defence most commonly given by psychologists and others in the cognitive testing profession (often called psychometricians). So if most of those deemed, by some independent test, to be highly intelligent give 107 as the answer to {128, 117, ?, 98, 90} then 107 is the only acceptable answer. There are two problems with this defence: regression and relativism.

Consider a particular test, say, T_1. If T_1 is deemed to be a good proxy measure of some attribute (such as intelligence) because its scores correlate strongly with the scores on some other test for that attribute (say T_2), then it is legitimate to ask how is it that we can be confident that T_2 is a good measure of intelligence. To answer that we know that T_2 is a good measure of intelligence because its scores correlate strongly with the scores on some other test for that attribute (say T_3) merely invites the same question, pushed back one level in a potential infinite regress. What makes T_3 a valid measure? If T_4, what makes it a valid answer? And so and so on. The regress has to stop somewhere if this defence is to proceed. And it can only stop at some robust, uncontroversial test or measure of intelligence (perhaps something akin to an axiom or definition). Let's call it T_7. Now if T_7 is obviously a better test of intelligence than T_1, why not simply rely on T_7 and discard T_1? To retort that T_1 is easier or less costly than T_7 will be of little comfort to those whose education or career has been thwarted or mis-channelled courtesy of an inferior intelligence-testing regime.

While on the subject on inter-test correlations, it is puzzling why designers of intelligence tests are keen to note that their test correlates well with other tests. (This is usually referred to as a measure of *concurrent validity*.) The fact that test T_3 correlates well with T_2 and T_2 correlates well with the supposedly uncontroversial T_7 doesn't mean that T_3 correlates well with T_7. Correlation comparisons are non-commutative. Consider, for example, the following three datasets: A = {5, 8, 7, 2, 8, 9, 4, 6, 6, 7}, B = {6, 9, 9, 3, 7, 7, 6, 8, 6, 4} and C = {5, 6, 7, 5, 7, 5, 5, 9, 5, 5,}. Imagine that A is a list of the results from a test deemed to the best measure of some attribute or other, and B and C are the results from some proxy tests alleged to be good measures of the same attribute. The correlation between B and A, and between C and B, is approximately 0.6 in both cases. You might think then that the correlation between C and A should be about the same. But it is just 0.24.[10] So the fact that a test has a good correlation

10. You can easily check these correlation values using the CORREL function in Microsoft Excel.

with, say, a Wechsler intelligence test tells us nothing about how well the test correlates with the test that gave the Wechsler test its validity (if it has validity).

Let's turn now to the second problem with basing correctness on the answers of those independently judged to be intelligent. Suppose that more and more people begin answering "any number" to number-sequence questions on intelligence tests. This is to be expected given how well-entrenched is the notion that intelligence is inextricably linked with an ability to provide correct solutions to problems. Psychologists will, presumably, observe this trend and seek to rediscover the answers intelligent people give to number-sequence questions (for it is only on this basis, they argue, that intelligence can be attributed or denied). It is very likely, then, that at some time psychologists will discover that more than the critical number of intelligent people—the majority, the top 40%, those in the top three percentiles, or however it is defined—are now answering "any number" rather than 107 when given {128, 117, ?, 98, 90} on an intelligence test. Suppose that this critical number is reached at time t_2. Suppose further that testee A takes an intelligence test at t_1—which precedes t_2—and takes the same test again at t_3 (which follows t_2). Let us suppose also that t_1 and t_3 are not too far apart and that A is of an age at which intelligence is fairly stable. If A answers each question identically on both occasions—including answering the number-sequence problems with "any number"—then A is, presumably, more intelligent at t_3 than at t_1. This is indeed odd, but what is odder is that A's level of intelligence has changed solely because of a change in the answers *other people* have given to test items. This extraordinary, even absurd, conclusion follows from an insistence that a person is intelligent solely if they do what other allegedly intelligent people do.

Psychological testing is a serious business. A person's future might be mapped out, and their self-esteem affected, by the result they achieve on such a test. Where a test item calls for a closed response—as in the case of a number-sequence question—testees rightfully expect to be rewarded for getting the answer correct, and this means that the only acceptable answer should be the correct answer. Furthermore, employers, schools and universities using tests to select only the ablest testees will be poorly served by tests that can rank those who give the correct answer below those who give the standard or expected answer. Neither of these desirable outcomes—rewarding correctness and selecting the ablest—is guaranteed by the use of number-sequence problems in such tests. Nor are they guaranteed by the comparative approach to intelligence testing (which, of course, exerts its levelling influence over all types of intelligence-test questions).

The solution in the case of number-sequence questions is not to revise them so that they might become better test items. That is not possible if, at the same time, such items are to retain their purpose as discriminators of numerical ability. For to revise such items so that the acceptable answer is always the correct answer means that every such item—whatever the numbers in the given sequence—would have the very same answer: "any number". And this answer, once known, ceases to be the product of numerical reasoning. The solution, then, can only be to discard these questions altogether from psychological tests.

*

Intelligence tests might be good at discriminating some attribute or other (such as likely school performance), but they also discriminate in another sense—the pejorative sense of implied unfairness. For a start, intelligence tests have led to some horrific classification errors. American psychologist Lee Cronbach—a fan of tests—observed that:

> "At times unsound testing or bad interpretations of tests have assigned children to the retarded category who did not belong there."[11]

11. L. J. Cronbach op. cit., pp. 4–5.

For a youngster to be labelled retarded—especially when they are not—would be felt as nothing short of life-shattering. For Cronbach, what saved the day was that at least "the majority of children [suspected of] retardation [were] saved [by a further] psychological examination".[12] Well that's cold comfort for the approximately 60,000 Americans who were forcibly sterilised between 1907 and 1968 based in part on their scores on intelligence tests.[13]

Carrie Beck is famous for having challenged the sterilisation laws in 1924, after having been declared retarded and committed to an institution. She lost and was forcibly sterilised, as was her sister Doris. Stephen Jay Gould, writing about Carrie and Doris, states:

> "Neither would be considered mentally deficient by today's standards … Can one measure the pain of a single dream unfulfilled, the hope of a defenseless woman snatched by public power in the name of an ideology advanced to purify a race?"[14]

Eugenics is a type of discrimination that is clearly immoral. Happily it is largely a practice of the past. But discrimination based on intelligence test scores is still rife, even if the outcomes are less horrific than forced sterilisation. No intelligence test correlates perfectly with whatever gold standard it is being compared with. For example, if a test is being used to predict school performance, the correlation would be perfect if the correlation coefficient was +1. This would mean that everyone who scored highly on the test did very well at school and everyone who did badly on the test did badly at school. (It would also be perfect if the correlation was –1, but now a high score on the test would correctly predict poor school performance and vice versa.) But, as we've noted, the correlation coefficient is not perfect. It is about 0.5. This means that some who do well on the test will do poorly at school and some who did poorly on the test will do well at school. Consider the scores in table 1 below.

Table 1 What might a correlation of 0.5 signify?

Test	School		Test Mean	School Mean
104	8		104.25	7.7
106	7		**Correlation**	
98	9		0.50051894	
87	8			
110	9			
120	8			
115	9			
104	8			
111	9			
90	6			
102	9			
110	10			
125	8			
118	9			
106	7			
90	4			
105	7			
110	5			
99	9			
75	5			

In the first column are the scores of 20 students who sat an intelligence test and in the second column is how well each subsequently performed at school (on a scale of 1 to 10). The mean of each set of scores is given at the bottom of the table. I have concocted these scores, but that doesn't matter in the argument that follows. What does matter is that the correlation between the two sets of scores is 0.5 (the same as is reported in the literature).

12. ibid., p. 5.
13. P. R. Reilly, *The Surgical Solution: A History of Involuntary Sterilization in the United States*, The Johns Hopkins University Press, Baltimore, 1991, p. 94.

Now compare each score against the mean for that particular set of scores. You will notice that in 6 cases a test score less than the mean corresponds to a school performance score that is greater than the mean, and in 4 cases a test score greater than the mean corresponds to a school performance score that is less than the mean. In other words, in 50% of cases, the intelligence test was a poor predictor of school performance. You might think this is only academic. But what if the test, being considered useful in predicting school performance, is used to distribute limited advantages (such as scholarships)? Is it fair that six out of the 20 students (30%) missed out on a scholarship despite having above average-academic potential? And is it an efficient use of scarce resources to give 4 students (20%) a scholarship when its value will only be squandered? Correlation is simply too blunt a measure to use in making life-changing decisions.

The use of intelligence tests to gauge school performance is scarily self-validating. If a child doesn't score well on such a test, it is likely that they will feel inferior to those peers who did better. Perhaps they will be put in a stream for the less-smart and get to learn less. Perhaps they will suffer class snobbery and maybe even bullying. Might all this be enough to dull their motivation? The overall outcome is poor school performance—exactly as the test predicted. But what if they hadn't been given the test and had received the same encouragement and education as their peers? It is not conceivable that their school performance would have been better?

So intelligence tests can adversely discriminate against a minority (namely those who are bright but do not score well on the tests). But they can also discriminate against the majority. Students can be coached to do better at intelligence tests. A spokesperson for EduTest—an educational assessment organisation in Australia that designs intelligence tests for use in scholarship filtering—is reported as saying "We are constantly editing tests … to stay ahead of coaching colleges".[15] This is an admission that you can coach a student to a better mark on a scholarship test. (Otherwise why bother trying to stay ahead of the coaches?) Hence parental wealth—and a willingness to provide your children with the best opportunities available—will in part influence who does well on an intelligence test and hence who gets a scholarship. Poor parents won't be able to afford coaches. Their children are thus less likely to get a scholarship. Awarding scholarships under a scheme whereby the children of the wealthy are likely to score higher than the children of the poor is a form of discrimination. In this case, it may be discrimination against the majority, for in many countries, the majority of families struggle just to maintain a decent standard of living. They do not have the disposable income necessary to pay for extra-classroom coaching for their children.

That coaching—and schooling in general—can affect IQ has been known for a while:

> "A striking demonstration of this effect appeared when schools in one Virginia county closed for several years in the 1960s to avoid integration [i.e., government-forced racial desegregation], leaving most black children with no formal education at all. Compared to controls, the intelligence-test scores of these children dropped by about 0.4 standard deviations (6 points) per missed year of school."[16]

If the removal of schooling can lower one's intelligence, it seems reasonable to assume that additional schooling (through coaching) is probably going to yield a higher intelligence score than might otherwise be the case. And experiments have indeed shown that coaching can give one a higher score on an intelligence test: up to 6 IQ points with coaching alone and 11 IQ points where the coaching includes practice tests.[17]

Discrimination, whatever its form, warrants attention. Some discrimination, as we noted earlier, is unavoidable. A one-armed person with no mind-controllable prosthesis should not have the same opportunities as an able-bodied person to become a commercial pilot. But here

14. Stephen Jay Gould, *The Mismeasure of Man*, Norton, New York, 2nd ed., 1996 , p. 335.
15. K. Marshall, "Fierce Contest for Edge in Education", *The Age*, 17 March 2014, p. 9.
16. U. Neisser op. cit., p. 87.

the reason for the discrimination is relevant. You need two arms to fly a plane. But a poor background does not limit one's ability to be pilot, a judge, a surgeon—indeed, any professional activity. It is simply not a relevant consideration. To limit one's access to a profession—and the rewards that go with it—on a basis of a tool that favours the wealthy is morally questionable, to say the least.

*

So what do intelligence tests measure if, as seems to be the case, they don't measure intelligence? Perhaps at best they measure nothing more than a person's ability to answer the type of questions that typically appear on intelligence tests. That is meant to be a criticism of intelligence tests. But some psychologists have gone so far as to define intelligence as what these tests measure: "If we agree to define intelligence as what the tests of intelligence tests, there is a good deal that we can say about it."[18]

Now definitions, obviously, are essential in any advance of knowledge. But do we always advance our knowledge by redefining concepts in ways that don't match common usage? It is true that sometimes common definitions do need to be rethought. Our concept of, say, *consciousness* needed redefining so as to exorcise it of the unscientific concept of *soul*. The concept of a *proton* as a fundamental building block of matter also had to be redefined as we discovered it to be a composite particle: a collection of quarks. And there are plenty more examples like that. But we do not advance knowledge by redefining concepts so as to make them unrecognisable as the things that spurred our curiosity in the first place. If we are curious about how birds fly, we do not redefine the word *bird* to mean aeroplane and divert the discussion to aeroplanes. We want to know about birds, and we know that birds are simply not aeroplanes (unless lexicographers have misled us for centuries). Likewise, I know that someone who answers a number-sequence problem with "any number" is more intelligent than someone who gives a single number. That is encapsulated in what we understand by intelligence. Define it in some other way if you like—test prowess, perhaps—but now we are not talking about the same thing. You have moved the discussion to other grounds. You have engaged in definitional fiat. Fine. Your argument might have some academic or armchair interest. But I still want to know about intelligence.

The standard intelligence test is primarily a test of one's numeric, verbal, logic and pattern-recognition ability. In addition to number-sequence and shape-sequence questions of the sort we've explored earlier, the tests ask questions like "Languages are to meaning as philology is to [erudition, philosophy, ethics, semantics or grammar]" and "At the end of a banquet 10 people shake hands with each other. How many handshakes will there be in total?"[19] It is not difficult to imagine that a highly gifted composer such as Stravinsky, a brilliant artist such as Monet or a revolutionary choreographer such as Balanchine would do poorly with such questions. Their minds are occupied with matters far less trivial than mind puzzles (and probably were so too during school years). And yet if intelligence is what is measured by intelligence tests, the likes of Stravinsky, Monet and Balanchine would probably be considered duffers. The absurdity of that is no doubt part of the reason for the widespread interest in Howard Gardner's theory of multiple intelligences. Gardner extends the notion of intelligence to include musicality, kinaesthetic ability, interpersonal skills and self-reflective aptitude to the range of skills tested by the standard intelligence test.[20] Such a view seems to more accord with our common understanding of intelligence. Paper-and-pen tests of intelligence simply can't measure these abilities, and other abilities recognised as markers of intelligence, such as giving an extemporaneous talk or being able to find one's way in a new town.

17. R. L. Bangert-Drowns, J. A. Kulik & C. C. Kulik, "Synthesis of Research on the Effects of Coaching for Aptitude and Admission Tests", *Educational Leadership*, vol. 41, iss. 4, December 1983/January 1984, pp. 80–82.
18. E. G. Boring, "Intelligence as the Tests Test it", *New Republic*, iss. 36, 1923, p. 37.

Perhaps intelligence cannot be defined. If two dozen psychologists can give two dozen different definitions of intelligence, then perhaps there is no definition, no single set of necessary and sufficient attributes that together indicate its presence. In his *Philosophical Investigations*, the Austrian philosopher Ludwig Wittgenstein noted that some words are nebulous. They are understood not as representing a discrete thing but as indicating that the thing in question shares certain attributes with something else that's called by the same name. The things in question share at most a family resemblance, not some set of necessary and sufficient attributes. An example is the word *game*. However we try to define it, some games escape our net. For example, we might start off thinking that games are played for fun or recreation, but this would rule out games like cricket and football when played professionally. Some games have scores, but others don't. Some have teams, and others don't. And so on. Nothing seems to encapsulate all games. Perhaps that is the case with the notion of intelligence. There is nothing in common with everything that we rightly call intelligent, merely numerous partial overlaps and resemblances. If that is so, then limiting intelligence to what intelligence tests measure is as nonsensical as limiting games to contents between two people.

In its zeal to quantify everything, science has got itself into an unholy wrangle. It has modelled intelligence on a set of analytical skills and forgotten that a model is a mere simplification. That would not be a problem if the thing being modelled was not a highly charged notion. But it is. Life's opportunities are dependent on how intelligent we are perceived. The wrangle—and the tarnish to science—began with the spurious claims to objectivity given by old-school behaviourists such as Eysenk and Skinner. It was disingenuous of them to claim objectivity for the research they did to show that blacks in America have lower intelligence scores than whites. When the whole family of meanings we associate with the word intelligence is taken into account, no doubt we could devise an intelligence that showed that blacks are more intelligent than white. Alas, the stench of racism was allowed to waft down the corridors of science.

A parallel with another highly charged yet nebulous concept will be a god way to end this paper. Consider the concept of beauty. Imagine that a scientist attempts to quantify beauty. Perhaps they develop a set of criteria which happens to correlate well with some other test of beauty (perhaps with the views of artists, or with statistics on marriage rates, or whatever). It's not a perfect correlation (say 0.5, much as the correlation between IQ test scores and school achievement). What this means is that the test will declare some people who are ugly to be beautiful and some who are beautiful to be ugly. Let's imagine that this test is put into service with as much alacrity as intelligence tests were. Companies start to use it to select employees or clients. From a batch of applicants, a TV station decides to recruit as cadet news reporters only those with the highest beauty scores. Retailers decide to refuse employment to anyone with a below average beauty score. A match-making company decides to accept as clients only those with high beauty scores. Lives could be ruined. There would, I suggest, be uproar as the unfairness of the test sunk in. Yet we put up with lives being ruined from the use of intelligence tests.

19. These questions were taken from a sample intelligence test published by IQ Test Labs. See http://www.intelligencetest.com/index.htm. (Viewed 10 March 2014)
20. M. Gardner, *Frames of Mind: The Theory of Multiple Intelligences*, Basic Books, Philadelphia, 1983.