

3. The Construction of Intelligence

Understanding that artificial reconstruction of the human environment serves as a fundamental component of the human transformation does bring us one step closer to identifying a cause for this uniquely human occurrence. We can now recognize that humans have somehow gained a broader perceptual awareness of their surrounding world, an awareness that has allowed the species to break free from a strictly survival-and-procreative focus, and this broader perceptual awareness, centered around the concepts of artificial pattern, structure and form, has helped spark an effective reshaping of the human environment and a radical expansion of human behavior. So we are nearly prepared to ask, what characteristic, what element within the human population, could be the key that has unlocked this broader perceptual awareness?

We will explore that question in great detail shortly, but for now I must ask your indulgence to postpone that discussion for a bit longer. The purpose of the delay is to take some time to examine in greater depth the activity of human environmental reconstruction. As it happens, we can approach this topic from an entirely different angle and can view it in a completely different light, one that has modern scientific consequence. This additional analysis will prove to be fruitful: it will provide greater information about what is the most basic human quality that lies at the core of the human transformation, and furthermore, this analysis will help solve a conundrum that has been puzzling scientists for at least the last thirty-five years.

As a reminder, in characterizing modern humans as both animal and construct, the word *construct* was used to denote two separate but related characteristics:

1. The artificial reconstruction of the human environment; and
2. The novel human behaviors engendered by that artificial reconstruction.

What I am going to propose now is that there is *another* word, besides *construct*, that can be used to capture the combination of these two related characteristics, a word that is in widespread use and that is clearly foundational to what makes humans unique. That other word is *intelligence*.

Although it might seem at first that I am employing the word *intelligence* in a somewhat different sense than that of the typical connotation, rest assured that there is nothing arbitrary about my proposal. In the discussion that follows, I will directly connect human environmental construction to the basic tool for measuring human intelligence, the IQ exam, and I will go on to demonstrate that performance on an IQ exam not only assesses the traits we commonly associate with intelligence, but also, and quite equivalently, assesses our interaction with the artificial and constructed material of the human world. This demonstration will directly link human intelligence to the observable features of the human transformation, and thus will mark intelligence as the most fundamental human quality underlying the human transformation (as opposed, for instance, to say language ability or collective learning). But this demonstration will also necessitate a reassessment of the word *intelligence*, requiring that intelligence be understood more broadly than as just a neurological phenomenon. We will see that to arrive at a complete understanding of intelligence, to provide a cogent explanation for intelligence scores, we must incorporate into our definition of intelligence an adequate provision for the totality of constructed artifacts existing within the human environment, the material out of which intelligence is constructed.

And there will be a bonus.

During the twentieth century, it was discovered that each successive generation was scoring progressively better on intelligence exams; in other words, raw intelligence scores were consistently and significantly increasing over time. This increase in performance was large enough that it required intelligence exams to be modified on a frequent basis, typically towards greater complexity and difficulty. Several researchers had noticed this phenomenon, but it was

James Flynn in the 1980s who provided abundant evidence that the phenomenon was essentially universal, thereby drawing greater attention to it, and the phenomenon would be eventually dubbed the Flynn effect. Over the past several decades there have been many attempts to explain the Flynn effect—increasing genetic robustness, better nutrition, expanded education, abstruse theories about fast and slow life, etc.—but none of these suggestions have proven to be compelling, and thus the Flynn effect has continued to remain a scientific mystery.

But as it happens, our subject from the previous section, the non-evolutionary and artificial environmental reconstruction that lies at the core of recent human history, turns out to have great relevance to the phenomenon of an increasing human intelligence. By linking the contents of an IQ exam to the observable features of the human transformation, and by exploring the potential for human intelligence performance throughout the course of human history, we will come to realize that an increase in human intelligence is not just a twentieth-century aberration, but is indeed a fundamental property of the human transformation itself. That is, the Flynn effect has been accompanying humanity ever since the beginning of the human turn, with the corresponding increase in human intelligence directly tied to the amount of artificial construction that has been accruing in the human environment, thus yielding an extremely straightforward and elegant explanation of the Flynn effect.

Intelligence is one of those words that does not seem to have a completely agreed-upon definition. There are several characteristics commonly associated with intelligence—reasoning ability, problem solving, capacity to learn, creativity, etc.—but none of these characteristics, singularly or in combination, quite manages to capture everything that is meant by the word *intelligence*. There appears to be a certain *je ne sais quoi* element to intelligence; we easily recognize smartness when we see it and experience it, but we seem to have a hard time putting our finger on exactly what it is.

Scientists sidestep this difficulty by sticking to what they can measure. The scientific approach to intelligence focuses both on the basic tool for assessing intelligence, the IQ exam, as well as on the statistical analyses that can be obtained from IQ exam performance. These quantitative efforts have been historically productive. The first IQ tests were developed around the year 1900, and throughout the following century, researchers kept themselves busy putting IQ scores through a series of statistical investigations, teasing out much of what is currently understood about human intelligence. One of the earliest and most basic results from these efforts came in the use of factor analysis to arrive at a *general* factor for intelligence, typically quantified in a statistic known as Spearman's *g* (named after Charles Spearman, its developer). In short, Spearman's *g* captures the notion of a general intelligence ability, one that shows up in the correlation of performance across different types of intelligence tasks—that is, a person who can score well on one type of intelligence test will also likely score well on the others. Furthermore, IQ exam analysis, in combination with identical twin and other family-based studies, has shown that one's general intelligence ability is influenced in large degree by one's genetic makeup—more so than, for instance, one's upbringing or environment—thus pointing to a significant biological/neurological component underlying individual intelligence differences.

These exam-driven conclusions famously prompted psychologist Edwin Boring in 1923 to define intelligence as the thing that gets measured by intelligence tests. Boring partly intended his aphorism as a warning, noting the extremely insular nature of many intelligence statistics. But he also seemed to suggest that in a certain sense his seemingly tautological definition was capturing a meaningful truth, one that could be built upon if approached in the right way. In my opinion, Boring's definition actually turns out to be both valid and quite good, that defining intelligence by linking it to intelligence exams actually has a large degree of substantive merit. I say this for two reasons, one of which is widely known, and the other of which seems to have been given hardly any attention at all.

The widely known reason is this. Although it is indeed tautological to say that intelligence is what gets measured by intelligence tests, it also has been repeatedly

demonstrated that intelligence tests have a significance that goes beyond that which they measure. This is because performance on intelligence exams has been shown to be highly correlated with many aspects of human life, aspects that have great importance to people navigating their way through the human world. Academic success, career prospects, socioeconomic circumstances, even health and longevity—all these areas correlate significantly to IQ exam performance, with better IQ exam performance pointing towards cheerier prospects in actual life. So whatever quality it is that is being measured by an IQ exam—call it intelligence if you will—it tends to translate to happier prospects outside the exam room, meaning that intelligence statistics serve more than just an inward-looking purpose.

The less well known reason centers around the question of what exactly is it about any particular test—that is, what observable features must that test have—in order for it to qualify as an *intelligence* test. Consider these two possibilities. In the first instance, I write down a dozen or so math and logic problems on a sheet of paper, give the paper to a subject to fill out, and then afterwards sum up the correct answers. In the second instance, I give the subject a sheet of paper and ask him or her to crumple it up and throw it as far as he or she can, and then afterwards I measure the distance the paper traveled. Nearly everyone would agree that the first instance could qualify as an intelligence test, and that the second instance would not. But why? Both examples are a test of human ability and both provide a numerical measure that can be compared, so why is one considered appropriate for assessing intelligence and the other one is not?

One way of answering this question might be to note that performance on the first test will likely have some correlation with other intelligence tests, and that the second test is unlikely to have such a correlation. But this just gets us back into tautology land, where a test's acceptance into the intelligence test club is done solely by reference to tests that have already been accepted. This is not really what we are after. What would like to be able to do is to identify a test as an intelligence test by something other than saying that it is used for measuring intelligence. In other words, if all intelligence tests were described simply as perceived artifacts, without reference to what purpose they serve, would there be a characteristic or characteristics they all would have in common?

This question prompts an investigation into the *contents* of an intelligence exam, which in my estimation is the most important thing *about* an intelligence exam, even if the topic seems to have been given very little attention in the literature. To understand why it is that an IQ exam succeeds in measuring intelligence, we need to better understand what it is that an IQ exam is composed of. This investigation also has a historical component to it: if we were to contemplate the possibility of intelligence exams having been available throughout human history, would the contents of those exams have remained static over time, or would they have needed to be altered? And this is where the investigation connects to the broader topic we have already been discussing, for as we shall see, the contents of an intelligence exam cannot be arbitrary, but instead must have a particular characteristic to them, a characteristic intimately connected to the fundamental features of the human transformation.

Some of the earliest IQ exams to gain acceptance and application were the Stanford-Binet and Wechsler scales, tests that are still in widespread use today. The contents of these particular tests cover a broad range of subjects, including vocabulary, arithmetic, number-letter sequencing, general knowledge, similarities, visual-spatial puzzles, and so on. The popularity of these exams has of course influenced what is considered to be the standard content for intelligence tests, but it should also be noted that there are many alternatives available. For instance, Raven's Progressive Matrices, a test composed of a logical series of spatial/temporal patterns, has been shown to be effective in assessing non-verbal intelligence, and has been one of the tests displaying the greatest amount of Flynn effect over the past century. Indeed there are essentially an endless number of tasks that could be successfully employed, at least to some degree, to assess various aspects of intelligence, everything from multiplication tables

to chess problems to driver's license exams. So what is it about the contents of all these tests and tasks that connect them to the subject of intelligence, what could be their common thread?

One approach to discovering the type of content that an intelligence exam must have would be to consider first those tests that are clearly *not* acceptable for assessing intelligence. Much like the example of throwing the crumpled up sheet of paper, tests that measure for qualities such as running speed, strength, visual acuity, agility, fertility and so on would not be acceptable as measures of intelligence, despite the fact these qualities are often important to humans in their everyday lives. The contents of such tests would have a particular nature that would appear to disqualify them—namely, these tests would be composed primarily of activities that are athletic, physical, biological, and we seem to intuitively understand that these particular abilities and activities are exactly the ones that need to be *excluded* when assessing intelligence. This intuition has a rational basis. When we examine the types of tests that are being ruled out here, and when we consider the activities and abilities that underlie them, we recognize that we could actually administer such tests—albeit with some difficulty—even to wild animals, and we also could have administered such tests to humans back when they were in the state of being pure animals themselves. The contents of such tests have an evolutionary/biological basis, they are intimately connected with survival-and-procreative fitness. Thus they measure the *animal* aspect of humankind, and they do not measure the aspect that has been added to humanity in the more recent years.

And this gets us to the common thread underlying intelligence tests and tasks.

When we examine the contents of any intelligence test, the one word that should immediately strike us is *artificial*. Artificial patterns, artificial structures, artificial form. Intelligence tests are composed of artificial constructions—words, digits, sequences, puzzles, matrices, etc.—and when we score an individual's performance on an intelligence test, what we are assessing is that person's ability to understand and manipulate these artificial constructions. And since *all* intelligence tasks are composed of artificial constructions, it should come as no surprise that performance across these tasks tends to correlate, with one's general intelligence ability essentially being a measure of one's overall effectiveness at manipulating artificial construction at large.

But artificial construction at large is exactly the defining observable feature of the human transformation, and human *responsiveness* to artificial construction is the telltale behavior distinguishing humans from all the other animals. The originators of the first intelligence tests may not have realized it, but what they were building were assessments of human modernity, assessments of human capacity for the non-natural and constructed artifacts that have been accruing in the surroundings ever since the beginning of the human turn. The artificial pattern, structure and form to be found on every intelligence test mirrors the artificial pattern, structure and form to be found in the human environment, and thus the contents of an intelligence test serve as a *proxy* for the type of artificial complexity that needs to be navigated by humans in their everyday lives. And so here too, it should come as no surprise that performance on these proxies correlates significantly to successful prospects in the everyday world, because for most humans today, effective navigation and manipulation of the environment's artificial features is often far more important than effective navigation in the biological/evolutionary realm.

A further approach to exploring IQ exam contents and their relationship to the artificial reconstruction of the human environment—that is, to the human transformation itself—would be to contemplate IQ exam performance over the entire course of human history. Of course since the first intelligence tests were not created until around the year 1900, we do not have explicit data from before that time, neither on what characteristics intelligence exams might have possessed during those earlier years, nor on the level of performance such exams might have evoked. Thus we will need to resort to some imagination. But even with this limitation, it

will take very little in the way of effort and reasoning to convince us that intelligence exams from those earlier times would have had to be crafted much differently than they are today, and that intelligence performance throughout those earlier years would have been much diminished from what we observe in the general population now.

To give some visualization to our investigation, let's assume that scientists have developed a means of time travel that allows them to go back to earlier points in human history for the purpose of administering intelligence tests. We can begin by going back around five thousand years, to the agrarian villages that would soon coalesce into the Mesopotamian civilization. Our time-traveling scientists will have brought with them the current iteration of the Stanford-Binet and Wechsler scales, with the hope of administering translated versions of these tests to the agrarian population. But of course the scientists will quickly encounter some major challenges. In the first place, there is no hope of administering a written version of these tests—five thousand years ago, writing was still on the verge of being invented—and so any intelligence task would have to be delivered orally, making many portions of modern IQ exams essentially impossible to administer. But the problem goes beyond just this. For instance, much of the vocabulary being tested would have to be adjusted, because a large portion of the vocabulary showing up on modern exams represents objects and concepts that have only come into existence after Mesopotamian times. Arithmetic too would have to be essentially removed, since numeracy beyond perhaps counting up to one-two-three was not yet widely practiced. Even something like Raven's Progressive Matrices, a non-verbal exam with geometric patterns that could be drawn perhaps on tree bark—or in the dirt, for that matter—would still cause problems for this particular population. Having been raised in a world without the regularity of clocks, street grids, etc., the Mesopotamians would likely find the rigidly structured patterns of the Raven's questions to be far outside their ken, with few having even an inkling of what was being asked of them.

Thus the scientists will soon find they need to abandon their efforts at administering any modern IQ exams they may have brought with them, and might even be wondering if Mesopotamian performance on such exams should be essentially marked near zero. But these scientists want to obtain an accurate and observed measure of the overall level of intelligence for this population, and also to assess whether there are any intelligence differences *within* the population, so the scientists scramble together an alternative test, a spoken and drawn exam composed out of tasks that the Mesopotamians can more easily relate to and understand. The vocabulary problems end up being more basic than what is found on modern tests, but at least a majority of the Mesopotamian test-takers can effectively respond. The mathematical and geometrical questions also will seem crude by modern standards, but at least they now correspond to what Mesopotamians encounter in their everyday lives. More questions can be composed out of the basic structural features arising from such activities as simple farming, irrigation, pottery making and so on. The resulting test might seem quite simple to the scientists—one of them might even remark, "Almost everyone from our era would score nearly 100%"—but when administered to the Mesopotamians, a familiar range of outcomes does emerge, confirming that these scientists have done an admirable job at putting together their makeshift exam. Some Mesopotamians score well, some score poorly, and most score somewhere in the middle. "Well, at least the *curve* looks familiar," one of the scientists remarks. The scientists might then leave these Mesopotamian times with a vague sense of puzzlement over what all these results might mean, but at least they have acquired some substantive data, and can continue onwards with their historical quest.

The next temporal destination goes back to around twenty-five thousand years ago, before the time of agriculture and villages but after the out-of-Africa migration. These humans are still universally hunter-gatherers, but they are not without some modern accompaniments: crafted weapons, clothing, fire pits, ornamental jewelry, cave paintings. Our scientists know better this time than to bother with the Stanford-Binet and Wechsler scales, but their first thought is to offer the makeshift exam they had crafted for the Mesopotamians to this new population, to see if that exam could still coax out some meaningful results. However, this turns

out to be mostly not the case, and this is because much of what the scientists had developed for the Mesopotamians was forged out of the agrarian concepts and abilities that had only been introduced into human experience sometime just before the Mesopotamian era, and thus would be utterly unfamiliar for these hunter-gatherer tribes. So once again our resourceful scientists patch together a different kind of test, one with tasks more familiar for their nomadic subjects. Out of necessity, the vocabulary is now even more limited than what was applied to the Mesopotamians, and the numeracy and geometric questions end up seeming almost toddler-like. Other questions are composed out of the underlying skills the scientists can glean from observed practices of fire making, weapon construction, animal rendering, and simple cooking. Indeed, so naturalistic and primitive would appear to be the lives of these hunter-gatherers that one of the scientists asks if they should not instead be testing for such skills as speed, strength and sexual dominance, until the other scientists object that these are not the kinds of skills one typically associates with intelligence. So the scientists stick with the basic exam they can manage to form out of the material at hand, and once again they find their efforts rewarded with a familiar set of results. Some of the hunter-gatherers score well on the new exam, some score poorly, and most score somewhere in the middle, demonstrating that for this population too there is an observable degree of intelligence capacity, with the capacity differing in the normal way from person to person.

Finally the scientists travel back to a few hundred thousand years ago, when humans had not yet begun to reconstruct their environment, when the surroundings remained almost entirely natural and humans were still essentially living the lives of pure animals. The scientists quickly realize, with regard to intelligence testing, that there is really nothing they can do. It is not just that the modern exams like Stanford-Binet and Wechsler cannot be applied, and it is not just that the exams developed for the Mesopotamians and previous hunter-gatherers cannot be administered as well, it is that these scientists cannot even craft a new kind of test, one more accessible to this ancient population. This is because there is nothing out of which a new test can be composed—no abstract language, no mathematics, no crafted weaponry, no clothing, no cooking. There is not a single artificial feature to be found in this environment, and there is not a single human behavior that cannot be described as being animal-like. The only abilities these scientists could possibly test for and measure would be the kinds of abilities that could just as easily be found in almost any other type of creature, abilities that are entirely physical and biological, abilities directly concerned with survival and procreation. Thus the scientists conclude that there is no intelligence to be found or measured in this population—at least no intelligence beyond that which could be attributed to *any* animal population.

But the scientists no sooner arrive at this conclusion than they begin to have some second thoughts. After all, the hunter-gatherers they had just previously visited, from twenty-five thousand years ago, did display some measurable intelligence capacity, and furthermore these ancient humans now standing before them, from a few hundred thousand years ago, do not appear to be, physically or genetically, all that much different from the previously visited populations—biologically speaking, all the visited humans have appeared to be essentially the same. So why did the previous hunter-gatherers display some level of intelligence, while this more ancient group seems to be displaying none at all? For that matter, why have *any* of the visited populations displayed a different level of intelligence from the others, or from modern norms?

“You know what I wish we could do,” says one of the scientists. “I wish we could take one of these newborns from this ancient population back with us to modern times, and then raise him there. What I would like to know is this: would he still display no intelligence at all, or as he grew up in the modern world would he begin to demonstrate a level of intelligence similar to other modern humans?” The scientists begin to think about this question, and they end up thinking about it for a very long time.

The conundrum of the scientist's question is this. On the one hand, If intelligence is simply a neural ability, then these ancient humans from a few hundred thousand years ago, because they are demonstrating no measurable level of intelligence, must be entirely lacking in this neural ability—an ancient newborn raised instead in the modern era would thus continue to display no signs of intelligence. Also, this neural ability must have somehow been introduced into the human population by the time of the hunter-gatherers of twenty-five thousand years ago, because that population was capable of taking an intelligence test, simple though it may have been, and of displaying a measurable level of intelligence. Furthermore this neural ability must have significantly increased by the time of the Mesopotamians, who were capable of taking a more complex intelligence exam, and thereby demonstrating a greater level of measurable intelligence. And finally, this neural ability must have increased by orders of magnitude yet once more by the time of modern humans, who find themselves perfectly capable of handling the complexities of Stanford-Binet and Wechsler. But on the other hand, there is no physical or genetic basis for such a large-scale, species-wide and progressive neural change—such an immense and ongoing biological alteration would utterly defy almost every evolutionary principle. The biological expectation is that neural ability would remain roughly the same throughout all the generations. But if neural ability has not changed for *Homo sapiens*, then what could be driving the human intelligence differences apparent across the entire course of human history?

If we were to spend enough time thinking about the scientist's question, we might eventually realize that there is every reason to expect that a newborn snatched from an ancient human population and raised in the modern era would achieve a level of intelligence commensurate with that of his modernly born peers. Furthermore, if that same newborn were instead to be raised among the hunter-gatherers from twenty-five thousand years ago, he would achieve a corresponding level of hunter-gatherer intelligence, and likewise if that newborn were to be raised among the Mesopotamians he would achieve a Mesopotamian level of intelligence. This newborn, along with his fellow humans from a few hundred thousand years ago, does not lack for intellectual ability—biologically speaking, he should possess the same neural capacity as do other sapiens. What he and his fellow ancient humans lack is any artificial complexity in their surrounding environment towards which they can *apply* their intellectual capacity. Human intelligence is not simply a neural ability. Human intelligence is instead the effective manipulation and navigation of the artificial construction to be found in the human environment, and thus no artificial construction translates to no measurable intelligence.

Furthermore, an *increase* in artificial construction translates to an *increase* in measurable intelligence. The quantity of artificial construction to be found in the hunter-gatherer environment of twenty-five thousand years ago was not nonexistent, but it was still quite small by modern standards, both in total amount and in depth of complexity. Any intelligence exam that would be appropriate and useful for that particular population would have to reflect this diminished quantity of artificial construction. A modern exam such as Stanford-Binet or Wechsler would be too complex and would contain too many features with which these hunter-gatherers would be unfamiliar, and thus would not serve as a useful proxy for hunter-gatherer intelligence. In comparable fashion, the quantity of artificial construction to be found in the agrarian Mesopotamian environment of about five thousand years ago was greater than that to be found in the hunter-gatherer environment, and yet still much less than that to be found in modern settings. Thus any intelligence exam intended for the Mesopotamians would need to be calibrated to reflect their particular quantity and type of artificial environmental construction. Stanford-Binet and Wechsler would still be too complex and unfamiliar, and thus would likely overwhelm the Mesopotamians, while at the same time the exam constructed for the hunter-gatherer population would not be challenging enough, leading to mostly excellent scores for nearly every Mesopotamian, so much so that individual intelligence differences could not emerge. Finally, humans of the early twenty-first century, having been raised in an environment overflowing with buildings, roadways, televisions,

computers and countless other complex artifacts, and having from a young age developed a broad vocabulary consistent with the broadness of their world, and having studied algebra and configured their phones and so on, would not find themselves challenged by exams constructed for the Mesopotamians and hunter-gatherers. Modern humans require an exam with greater complexity and greater variety in order to demonstrate their impressively broad skills at navigating the complex features of their modern world.

Of course all this has a direct connection to the Flynn effect. Intelligence exams with contents that contain greater complexity and greater variety are palpable indicators of an increased level of overall intelligence. And thus we can conclude that human intelligence has been increasing throughout human history, ever since the beginning of the human transformation, an increase that would have been reflected in the increasing complexity and difficulty of intelligence exams, had such exams been available to earlier populations. Furthermore, this increase in intelligence is precisely attributable to the fact that humans have been forced to navigate an increasing amount of artificial complexity in their surrounding world, meaning that the accruing amount of artificial environmental construction arising from the human transformation serves as the sole driver and the sole explanation of the Flynn effect.

Note that although our time-traveling scenario does require some imaginative reasoning, this reasoning remains entirely consistent with what has actually taken place over the past century, with respect to actual intelligence performance on actual intelligence exams. Over the past one hundred years or so, ever since the invention of IQ tests, each generation has scored better and better as the tests have aged, so much so that individual intelligence differences begin to disappear, rendering older exams essentially obsolete. Intelligence researchers have realized that they need to frequently modify their exams, adding new types of material and making the questions generally more complex and difficult. These researchers may have been somewhat puzzled as to what could be driving these changes, but as our historical analysis has indicated, these researchers have not been witnessing just a twentieth-century fluke. This increase in overall human intelligence, now dubbed the Flynn effect, has been shadowing human existence ever since the beginning of the human turn, and what these researchers have been witnessing is just the modern continuation of an extremely long-running phenomenon.

And from the historical perspective, there really should be nothing surprising about this phenomenon. If humans were indeed once pure animals, then it is perfectly reasonable to say that humans at that time displayed no intelligence at all, at least no intelligence beyond that which could be attributed to any animal species. And since humans today clearly do possess an observable and measurable level of intelligence, then over the course of human history, human intelligence has increased from absolute zero to a substantive number, and *by definition* that increase is a Flynn effect. The only question is whether that change has been sudden or gradual, but any reasonable reflection on the events of the human transformation will make it more than clear that the change cannot have been sudden, but that the increase in human intelligence has been progressing steadily throughout human history, an increase driven by the accruing amount of artificial construction contained within the human environment.

What I am going to do now is build a quantified scenario that incorporates the concepts and assumptions from the above discussion. The purpose of this scenario is to demonstrate the pattern of IQ scores that emerges when we take into account artificial environmental construction when developing, administering and scoring IQ exams. The pattern of IQ scores that emerges from this scenario will be exactly the same pattern as has been observed on actual intelligence tests over the past century, and this pattern will consist of both a stable level of general intelligence ability (such as that quantified by Spearman's *g*) as well as an increase in overall measured intelligence (that is, a Flynn effect). The scenario itself will make clear that

there is nothing contradictory about these two results. A stable general intelligence ability is perfectly consistent with an increase in overall measured intelligence, as long as our definition of intelligence incorporates the impact of the total amount of artificial construction contained within the human environment.

The scenario will be constructed out of the following assumptions and stipulations:

1. We assume we can administer IQ tests at four sequential points in time, call them Time 0, Time 1, Time 2, and Time 3. These points in time are of course arbitrary, marked out solely for the purpose of the demonstration, but if it helps to visualize them, we can think of Time 0 as being a few hundred thousand years ago, when humans were still pure animals, Time 1 as being around twenty-five thousand years ago, the era of efficient hunter-gatherers, Time 2 as being around five thousands year ago, the agrarian dawn of the Mesopotamian civilization, and Time 3 as being now, early in the twenty-first century. Do keep in mind, however, that this is just for visualization—the demonstration remains essentially the same for *any* four sequential points in time, as long as the interval between them allows for at least a new generation or two.
2. We assume we can measure and quantify the amount of artificial construction in the human environment at each point in time. Furthermore, we assume—consistent with human history—that the amount of artificial construction in the human environment increases with time. For this particular scenario, we will stipulate that the quantified amount of artificial construction measured at Time 0 is 0, that the amount of artificial construction at Time 1 is 10, that the amount of artificial construction at Time 2 is 20, and that the amount of artificial construction at Time 3 is 100. Per the visualization, this would be the equivalent of saying that there was no artificial construction in the human environment a few hundred thousand years ago, that there was a quantified level of artificial construction measuring 10 for the hunter-gatherers (a reflection of fire pits, simple cooking, ornamental jewelry, etc.), that there was a quantified level of artificial construction measuring 20 for the Mesopotamians (a reflection of simple abodes, irrigation techniques, pottery making, etc.), and that there is a quantified level of artificial construction measuring 100 in the early twenty-first century (reflecting highways, computers, skyscrapers, etc.), a scale that likely shortchanges the modern era, but still serves well for illustrative purposes.
3. We assume that at Times 1, 2 and 3, we can construct and administer a battery of IQ tests for the population of that era. Furthermore, we assume that each of these battery of tests is an equally accurate proxy for the amount and type of artificial construction contained within the environment at that time. Thus, the IQ tests of Time 1 have contents that reflect the measured level of 10 for artificial construction at Time 1, the IQ tests of Time 2 have contents that reflect the measured level of 20 for artificial construction at Time 2, and the IQ tests of Time 3 have contents that reflect the measured level of 100 for artificial construction at Time 3. In practice, this means that the tests become more varied and more complex as time goes on, just as was described for the hunter-gatherers, Mesopotamians, etc., and just as has been observed on actual intelligence tests over the past one hundred years.
4. Since the measured level of artificial construction at Time 0 is 0, we stipulate that no IQ test can be constructed and administered to this population.
5. Since the battery of IQ tests at each point in time is assumed to be an equally accurate proxy of that era's artificial environmental construction, we further assume that each population will score similarly on its particular set of exams. Consistent with this assumption, we stipulate that each population's results can be demonstrated through

the scores of three representative test-takers: one, Test-taker Low, who answers 50% of the test questions correctly, one standard deviation below the population mean; two, Test-taker Middle, who answers 60% of the test questions correctly, right at the population mean; and three, Test-taker High, who answers 70% of the questions correctly, one standard deviation above the population mean. We further stipulate that these same levels of representative test taking ability are in effect at Time 0, even though there is no exam to administer to the Time 0 population.

With these assumptions and stipulations in place, we can now describe what would happen over the course of the scenario, and what pattern of IQ scores would emerge from these populations.

There is not much to say about Time 0. With no artificial construction in the environment and no IQ exam to administer, all intelligence scores would have to be marked as zero, and this despite the fact that we have assumed the population does possess an intelligence capacity similar to that of the other populations. But in this particular environment, there is nothing towards which this population can *apply* its intelligence ability.

At Time 1, we administer the Time 1 battery of IQ tests to the population, and afterwards we norm the results by determining the statistical mean and standard deviation, just as is done with modern exams. With this accomplished, intelligence researchers can then perform most of the population analyses that they do for modern exams, since relative performance is all that is needed. General intelligence statistics can be determined, correlations with life outcomes can be scrutinized, and family studies can be conducted showing the extent to which individual general intelligence abilities are mostly genetically determined. *Relative* IQ performance within a population is the backbone for nearly all intelligence research.

But also at Time 1, we can do something that has not been attempted in modern studies. Because we have measured and quantified the amount of artificial construction in the Time 1 environment, and because we have developed a battery of IQ tests that reflects this level of artificial construction, we can use this information to determine an *absolute* intelligence score for each member of the population. It works like this. Test-taker Low answered 50% of the test questions correctly. Since the contents of the Time 1 tests reflect an amount of 10 for artificial construction in the Time 1 environment, we can say that Test-taker Low is showing an ability to master an amount of artificial construction equivalent to 5 ($50\% \times 10$), and thus we can assign Test-taker Low an absolute intelligence score of 5. Similarly, Test-taker Middle, by answering 60% of the test questions correctly, is demonstrating an ability to master a level of artificial construction equivalent to 6 ($60\% \times 10$), and can be assigned 6 as an absolute intelligence score. And finally, Test-taker High, by answering 70% of the test questions correctly, is demonstrating an ability to master an amount of artificial construction equivalent to 7 ($70\% \times 10$) and can be assigned 7 as an absolute intelligence score.

We could make use of these absolute intelligence scores when doing the population analyses of individual intelligence differences and general intelligence abilities, but of course the conclusions would turn out to be exactly the same, since these analyses hinge on the *relative* performance of the members of the population—and whether we start with normed results or absolute intelligence scores, the *relative* statistics will remain exactly the same. Thus absolute intelligence scores do not seem to add any further information when considering the intelligence characteristics of the Time 1 population, but these scores will turn out to be quite important when comparing intelligence performance against the other populations. For instance, we already know that the members of the Time 1 population are universally demonstrating a greater level of measurable intelligence than the members of the Time 0 population, for whom all absolute intelligence scores would have to be marked as zero.

At Time 2, we repeat the entire procedure: we administer the Time 2 battery of tests, we norm the results, we do the population analyses, etc. And many of the conclusions turn out to be the same as those determined for the Time 1 population, especially conclusions related to

individual intelligence differences, correlation with life results, and the impact of genetic factors in determining general intelligence ability. Thus the intelligence researchers could hardly be blamed for thinking that overall intelligence characteristics have not changed much in going from Time 1 to Time 2, so similar are all their studies and conclusions—studies and conclusions based upon *relative* exam performance *within* the Time 2 population.

But in fact Time 2 does have a significant difference. This difference shows up in two different ways. First, reflecting the increase in artificial environmental construction, the contents of the Time 2 battery of tests are more varied and more complex than the contents of the Time 1 battery of tests, but despite this extra challenge, the Time 2 population has still achieved the same level of raw performance on its particular tests. Second, the absolute intelligence scores of the Time 2 population are greater than those of the Time 1 population, doubling them across the board. For instance, at Time 2, Test-taker Low answers 50% of the test questions correctly, and since the exam reflects a level of artificial environmental construction measured at 20, Test-taker Low's absolute intelligence score is calculated to be 10 ($50\% \times 20$), twice that of the corresponding test-taker from Time 1. Time 2's Test-taker Middle receives an absolute intelligence score of 12 ($60\% \times 20$), and Time 2's Test-taker High receives an absolute intelligence score of 14 ($70\% \times 20$), in each case twice that of the corresponding test-taker from Time 1. Thus in an observable and measurable sense, the Time 2 population is displaying an overall level of intelligence that is twice that of the Time 1 population.

The story is the same at Time 3. The Time 3 battery of tests are administered, normed and analyzed, and once again, the researchers find that many intelligence characteristics of the Time 3 population are the same as those seen in the other populations, especially those characteristics connected to individual intelligence differences and general intelligence ability. This gives the appearance that intelligence is in some sense stable across time. But the differences at Time 3 are also noted, especially since the Time 3 differences are stark. The Time 3 tests, reflecting a five-fold increase in artificial environmental construction from Time 2, are now much more varied and more complex. And yet despite this additional variety and complexity, the Time 3 population still manages to achieve the same level of raw performance on its particular exams. This leads to the other stark difference, a massive leap in absolute intelligence scores. With the battery of tests now reflecting a level of artificial construction measuring 100, Test-taker Low at Time 3, answering 50% of the test questions correctly, achieves an absolute intelligence score of 50 ($50\% \times 100$). In similar fashion, Test-taker Middle at Time 3 achieves an absolute intelligence score of 60 ($60\% \times 100$), and Test-taker High at Time 3 achieves an absolute intelligence score of 70 ($70\% \times 100$). These absolute intelligence scores from Time 3 are considerably greater than all those achieved by the previous populations, suggesting that in an observable and measurable sense, intelligence is not stable at all but is instead significantly increasing over time.

The results of this scenario are summarized in the following table:

Sequential Point in Time	Test-taker	Raw Score	Normed Result	Absolute Intelligence Score (Artificial Construction x Raw Score)
Time 0 Artificial Construction 0	Low	50%	- 1.0 Std Dev	0
	Middle	60%	Mean	0
	High	70%	+ 1.0 Std Dev	0
Time 1 Artificial Construction 10	Low	50%	- 1.0 Std Dev	5
	Middle	60%	Mean	6
	High	70%	+ 1.0 Std Dev	7
Time 2 Artificial Construction 20	Low	50%	- 1.0 Std Dev	10
	Middle	60%	Mean	12
	High	70%	+ 1.0 Std Dev	14
Time 3 Artificial Construction 100	Low	50%	- 1.0 Std Dev	50
	Middle	60%	Mean	60
	High	70%	+ 1.0 Std Dev	70

The question begging to be asked is this: is human intelligence stable over time or is it instead increasing? But when we examine the circumstances of the scenario, we recognize that the question itself is ambiguous, and betrays a misunderstanding of what human intelligence is. Intelligence is commonly thought of as just a neural ability, but in fact neural ability is only half of the story. In order to understand and to explain the pattern of IQ performance over human history—both as outlined in the scenario and as seen on actual intelligence tests over the past one hundred years—we must expand our definition of intelligence. Measurable human intelligence actually consists of the interaction of two orthogonal factors: one, individual intelligence capacity (that is, neural ability); and two, the total amount of artificial construction contained within the human environment, the target towards which individual intelligence capacity can be applied. Over the course of human history, one of these factors has remained almost entirely stable, while the other factor has been significantly—indeed almost dramatically—increasing.

Thus when we ask whether human intelligence is stable or increasing, we need to clarify first exactly what it is that we are asking. If we are asking is *general intelligence ability* stable or increasing, then the answer to that question is that general intelligence ability is stable over time. This is exactly as we might expect for an ability that is being driven primarily by biological, genetic, and neural factors; there is little reason to think that *Homo sapiens* would have undergone a major change in this biological capacity since the earliest days of the species. Thus the first of the *Homo sapiens* would have had roughly the same intellectual ability as did the later out-of-Africa hunter-gatherers, who in turn would have had roughly the same intellectual ability as did the Mesopotamians, who in turn would have had roughly the same intellectual ability as do modern day humans. This stability shows up in intelligence

research, which gives no indication that general intelligence ability has been undergoing any kind of transition over the last one hundred years.

But if we are instead asking is the *measured level of overall intelligence* stable or increasing, then the answer to that question is that the measured level of overall intelligence is increasing over time. This is because the measured level of overall intelligence is determined by more than just general intelligence ability. To *demonstrate* intelligence, general intelligence ability must be *applied* to something, and that something is the artificial construction contained within the environment (or its proxy, the contents of an IQ exam). Furthermore, the greater the amount of environmental artificial construction, the greater the amount of intelligence that can be demonstrated. This was seen throughout the last century, a century in which televisions, computers, airplanes and so much more were being introduced into the human environment, an additional amount of artificial complexity that both was being navigated and mastered by humans in their everyday lives, and also was driving up raw performance on intelligence exams. Ultimately, it is the accruing amount of artificial construction contained within the human environment that determines the overall level of human intelligence, and it is the accruing amount of artificial construction contained within the human environment that drives the Flynn effect.

It has become something of a fad in recent years to suggest that the Flynn effect has now plateaued or even reversed. The studies on this matter have been somewhat conflicting, with some data suggesting that raw intelligence scores are no longer increasing and other data suggesting just the opposite, and of course it should be noted that over a short period of time there is bound to be some statistical noise. But one thing should be certain to us: based upon the above analysis and discussion, and based upon an understanding of human history, there is absolutely no reason to expect that the Flynn effect has ended or is reversing. Barring a catastrophe (such as civilization collapse), the human environment will continue to accrue greater and greater amounts of artificial construction, and the future generations will be obliged to navigate and to master this increase in artificial construction, and will thereby demonstrate an increased level of measurable intelligence.

One possible explanation for why the Flynn effect has become more difficult to detect in recent years is that the current IQ tests might be failing to keep up with the types of changes that are currently taking place in the human environment. In order for the contents of an IQ test to serve as an accurate proxy for the kind of artificial complexity humans encounter in their everyday lives, those contents must be modified over time to reflect the changing environmental circumstances. The popularity of tests such as Stanford-Binet and Wechsler has perhaps biased researchers on what should be the standard type of content for an IQ test, leaving them somewhat blinded to the necessary alternatives. For instance, we already have an example in the Raven's Progressive Matrices, an alternative kind of test focused on geometric and logical pattern, a type of artificiality that increased greatly in the technically and visually enriched twentieth-century world, and indeed Raven's Progressive Matrices has proven to be not only a good measure of general intelligence ability, it has also experienced some of the greatest Flynn effect movement throughout the last century. And even the so-called standard content of IQ tests betrays that there is really no such thing as a standard when it comes to the type of content that belongs on an intelligence exam. Arithmetic, for instance, might be considered a standard subject now, but there was a time not that long ago in human history when arithmetic did not even exist in the human world. By necessity, the contents of IQ exams must be fluid over time, because the most important requirement for the contents of IQ exams is that they reflect the amount and type of artificial complexity that is contained within the human environment, something that is often rapidly changing over time.

I can think of two recent changes in human complexity and artificiality that might not be getting adequately represented on current IQ exams. First, there is the notion of multitasking. When I was young, concentrating on one task at a time was considered generally the norm—

indeed it was often encouraged—but today, a great deal of human work and play is accomplished by doing a multitude of tasks in parallel. Think of all the office workers who can maintain several open windows on their computer screen, moving seamlessly from chats to emails to spreadsheets, etc. Or think of the teenagers who can engage with multiple applications on their smartphones while maintaining an ongoing conversation with friends. The younger generations, having grown up in a world more suffused with these multitasking demands, generally find such efforts to be easier to accomplish than do the older generations, for whom multitasking represents something of a major change. Multitasking has become an essential skill for navigating the parallel complexity of the modern world, and if this skill were more directly measured on IQ tests, the results would likely reveal not only that multitasking has become an important aspect of modern human intelligence, but also that there is a significant difference in performance between the younger and older generations.

A second area of change in recent environmental complexity is that humans do not interact *directly* with the artificial features of their surrounding world as frequently as they once did, but instead provide instructions to some type of machine which can handle the interaction for them (often with greater efficiency and accuracy). For instance, humans these days seldom perform arithmetic by hand, as they once commonly did; instead they program a calculator or computer to accomplish the task for them. This increases the amount of artificial complexity that can be navigated, because a correctly programmed machine can have a massive leveraging effect. But this also means that humans, to be effective in the current world, must acquire a new set of skills, skills falling under the general heading of coding or programming. And once again, it is the younger generations, having been born into a world that more frequently requires these types of machine instruction (just take a look at the current line-up of children's toys) that find such coding tasks to be easier to accomplish than do the older generations, who are in many ways still scrambling to catch up. If machine coding skills were given more attention on the current IQ tests, then the results would likely indicate, much like with multitasking skills, that machine coding is now a critical aspect of human intelligence, and that there is a significant difference in performance between the younger and older generations.

Whether it is through recognition that human intelligence has been increasing in areas not yet commonly measured, or whether it simply takes the passage of enough time, researchers will eventually realize that the Flynn effect is still very much with us here in the early twenty-first century, just as it has always been, and the Flynn effect will continue to shadow humanity throughout the coming years. The Flynn effect is not just a twentieth-century aberration. The increase in human intelligence is instead one of the most fundamental components of the human transformation itself, encompassing the two essential features of the human transformation, the artificial reconstruction of the human environment, and the human behavioral responsiveness to that artificial reconstruction. Thus in a very real and observable sense, human intelligence is *built*—it exists less inside our heads than it does in the surrounding environment. The artificial construction we now observe so abundantly all around us, this is the physical manifestation of intelligence itself, and since the amount and complexity of that artificial construction continues to increase over time, so does human intelligence.