

CHAPTER 2

Female Intuition Versus Male Reason *The Battle for Intelligence*^{*}

Her philosophy is not to reason, but to sense.

Immanuel Kant (1764)

Her logical thought is slower, but her associations quicker than those of man, she is less troubled by inconsistencies, and has less patience with the analysis involved in science and invention.

G. Stanley Hall (1904)

Immanuel Kant's conviction that women's nature is sense rather than reason surprised few scholars during the Enlightenment. Learned ladies, Kant believed, were worse than useless, and the very thought of women intellectuals interested in Greek philosophy or the foundations of mechanics seemed almost comical in his eyes.¹ Kant stood in a long and tenacious tradition of scholars convinced that the mind of a woman differs from that of a man. It can be traced back to Aristotle's influential contention that "the female is softer in disposition, is more mischievous, less simple, more impulsive, and more attentive to the nurture of the young; the male, on the other hand, is more spirited, more savage, more simple and less cunning . . . She is, furthermore, more prone to despondency and less hopeful than the man, more void of shame, more false of speech, more deceptive, and of more retentive memory."² At the beginning of the 20th century, psychology reiterated the idea that women are qualitatively different. The founder and first president of the American Psychological Association, G. Stanley Hall, held that women are intuitive and emotional, slow in logical thought, better at mental reproduction than production, and too impatient for analysis and science:³

^{*} This chapter is a slightly revised and shortened version of Gigerenzer (2022c).

¹ Kant (1764/2011). ² Aristotle (350 BCE/1984), pp. 948–949.

³ Hall (1904/1976), pp. 565, 651.

She works by intuition and feeling; fear, anger, pity, love, and most of the emotions have a wider range and greater intensity. If she abandons her natural naiveté and takes up the burden of guiding and accounting for her life by consciousness, she is likely to lose more than she gains, according to the old saw that she who deliberates is lost.

Hall, then president of Clark University, consequently opted against coeducation. Like Clark University, Harvard set up a female institution in the 1890s, Radcliffe College, next to the all-male Harvard College. But even there, women were treated differently. Not until 1967 did Harvard's Lamont Library open its doors to female students, an opening vehemently opposed by the administration and the majority of male undergraduates, on the grounds that females would distract male students and that there weren't even bathroom facilities for women.⁴ It took yet another 10 years before Harvard terminated its policy of admitting only one female student for every four male students.

Hall expressed what psychologists at the time held to be a law of nature, traces of which can be found in people's thinking today. When my colleagues and I asked representative samples of 21st-century Germans and Spaniards about gender differences, the result was surprising – or perhaps not. The vast majority of women and men, young and old, believed that women had better intuitions than men about personal affairs, but not about science and finance.⁵ And the rejection of learned ladies persists: Most contemporary American men in search of a partner on online dating sites find women with a Master's or PhD degree unattractive and prefer those with lower education.⁶

This chapter is a case study on how the lack of theory about the nature of intelligence enabled cultural biases about women to be presented as science by major psychologists. A discipline that is unaware of the errors in its history is potentially hazardous: "Those who cannot remember the past are condemned to repeat it."⁷ I reconstruct the history of the idea of a peculiarly female intelligence in three overlapping views. In the first view, from Aristotle through to the mid-19th century, the idea of intelligence as we encounter it today – as a general ability that is measurable and is largely independent of personality and moral character – did not exist. Instead, the difference between men and women was understood in terms of polarities that were a mixture of intellect, personality, and moral character, such as men's reason versus women's intuition. The notion of these

⁴ Masters (1986).

⁵ Gigerenzer et al. (2014).

⁶ Bruch & Newman (2018).

⁷ Santayana (1905).

polarities wore away in the mid-19th century and was supplanted by the concept of an inherited “natural ability” (soon to be named *intelligence*), mainly through the writings of the English polymath Francis Galton. As a consequence, in this second view, men and women no longer differed in quality, but in quantity: On average, it was thought, women had inherited a smaller share of intelligence. The psychologist Lewis Terman put an end to this view by eliminating particular test items from his test (the Stanford-Binet test, which I discuss in the section “Binet’s Intelligence Test Crosses the Atlantic and Becomes Seen as a Test of Genetic Ability”) and balancing the rest so that girls and boys had the same mean IQ. The eliminated items landed in a personality scale called “masculinity-femininity” (see the section “How Differences in Intelligence Become Differences in Personality”), which illustrates the arbitrariness of what counted as a measurement of intelligence rather than of personality. What remains debated to the present day is the third view, promoted by sexologist Havelock Ellis (1859–1939). It alleges that men’s intelligence varies more than that of women, implying the existence of more male idiots and geniuses.

Whatever the hallmark of a peculiarly female intelligence has been – polarities, lower average, or lower variability – it has served the dubious purpose of justifying men’s dominant role in society. Similarly, whatever the supposed mental differences were, these became presented as part of the natural order, expressed in the female body and women’s reproductive function.⁸

Before Intelligence: Female Intuition Versus Male Reason

Intelligence, as we know it from IQ tests, refers to a general ability that can be measured by a single number and is assumed to be largely independent of personality and moral character. IQ tests have been given to millions of children, recruits, and job applicants, and continue to influence access to education and jobs. The IQ has often been presented as a hard fact, and debates have raged over how much of its variability is due to nature or nurture. These debates ignored the fact that intelligence as we know it was “invented” in the late 19th and early 20th centuries.

Before that time, what we now call the intellect was considered neither as a single general ability nor as largely unrelated to moral and personality traits. Rather, psychological theories conceived the mind as a collection of

⁸ Daston (1992).

faculties or talents. For instance, the key concept of *sensibility* in early 18th-century psychology encompassed both perceptual and emotional sensitivity as the precondition for empirical knowledge and the emotions of charity and compassion.⁹ Reason was even more closely identified with morality because the light of reason enabled one to recognize all forms of truth, including the distinction between good and evil. No single one of these faculties or a combination thereof corresponds to the contemporary concept of intelligence.¹⁰

The prototypical male and female occupied opposite poles on the spectrum of these faculties. For instance, men were characterized by judgment, abstract thought, and genius, while women were considered to lack these and, instead, excel in intuition, concrete thought, and retentive memory. Male strength was opposed to female delicacy or bodily and mental weakness. This supposed weakness was, in turn, seen as evidence that nature intended women to confine themselves to the home and subordinate themselves to men. It was reasoned that because men's thought was abstract, they could comprehend truth, including moral truth, while women's concrete thinking prevented them from grasping abstract moral principles. Hence, women who lied or stole were considered incapable of understanding that their actions were evil. When Hall, in 1904, wrote that women, guided by intuition and feeling, were unfit for science and invention because they lacked patience, he was simply reiterating the timeworn conviction that women did not have the necessary self-discipline and stamina to reason by following a lengthy chain of argument.

Women's and men's virtues were also seen as diametrically opposed. For centuries (and in many contexts even today), chastity was considered the chief female virtue and its violation a cardinal sin for women alone. Timidity, in contrast, was a cardinal sin for men, but easily excused in women.¹¹ The view that women's intellect, character, and moral traits are intimately connected to their biology survived in various forms into early 20th-century philosophy. Consider the controversial Austrian philosopher Otto Weininger, hailed by Freud and Wittgenstein as a great genius.¹² In his book *Sex & Character*, Weininger drew on a wide range of philosophers and psychologists to assert that reasoning and feeling are equivalent in women who, as a consequence, are prone to suggestibility, hypnosis, and hysteria, as documented by Freud. These alleged flaws correspond to Aristotle's view that women's memory is easier to imprint. From biologists Sir Patrick Geddes and John Arthur Thomson, Weininger borrowed the

⁹ Rifkin (2002).

¹⁰ Daston (1992).

¹¹ Ibid.

¹² Dury (1984); Weininger (1906).

conviction that each cell in a woman's body is sexually marked to make the female in every respect passive, submissive, and lacking in personality.¹³ Unlike man, he wrote, "woman is non-logical and non-moral."¹⁴ Faced with the fact that more men stand trial for crimes, he argued that behind every lawbreaker is a woman who proposes the crime and profits from it. Weininger gained great popularity when he killed himself at the age of 23 at a spectacular site, the room in which Ludwig van Beethoven had died. This dramatic finale led to huge book sales and an enthusiastic reception by many contemporaries, including the Swedish playwright and novelist August Strindberg, who claimed that Weininger's book had finally solved "the problem of women."¹⁵

In sum, for millennia, a fairly consistent view reigned about women's intellect as differing fundamentally from that of men. My brief account does scant justice to the variations of this view among scholars and centuries. Yet, the common denominator between them is that there was no concept of a general intelligence, which was, instead, defined by a number of diametrically opposed polarities attributed to the prototypical male and female, a combination of what were only later separated into intelligence, personality, and moral traits. While the specific descriptions of the female and male attributes varied, they revolved around the poles of female intuition and male reason. This opposition, allegedly handed down by nature, was so manifestly true to psychologists and philosophers that evidence could not compromise the narrative. It was driven by motives outside of science, in particular the attempt to defend male's intellectual and moral dominance over females as part of the natural order. That is history, one might contend, and will not repeat itself. Yet, Chapter 3 shows that, in the 21st century, the opposition between intuition and reason has been resurrected in psychology and once again embraced as a persuasive narrative that is held to be self-evident.

The Invention of a General Inherited Intelligence

The idea of mental faculties was slowly abandoned in the late 19th century for that of a single overarching intelligence. However, the associated idea that this intelligence combines cognitive abilities, personality, and moral traits did not fade away until the early 20th century. The transition from multiple mental faculties to a single intelligence was driven not by data or by experiment, but by concerns outside the realm of science, chief among

¹³ Geddes & Thomson (1890).

¹⁴ Weininger (1906), p. 297.

¹⁵ Abrahamsen (1946).

them being Francis Galton's interpretation of evolutionary theory, his fascination with measurement, and his involvement with the fateful eugenics program.

Women Are Granted the Same Kind of Intelligence As Men, but Less of It

Galton, a cousin of Charles Darwin, promoted a strict distinction between nature and nurture, which had not been considered mutually exclusive before his time.¹⁶ This artificial distinction later led to a flood of psychological research seeking an answer to the (wrong) question of what percentage of the variation in intelligence is due to nature and to nurture (as opposed to asking how genes and environment interact, as in epigenetics). For Darwin's theory of evolution to work, it was clear that something must be passed on to the next generation and inherited by both boys and girls. In *Hereditary Genius*, Galton called this something *natural ability* (later known as *intelligence*).¹⁷ As he saw it, evolution implied that men and women must have the same kind of natural ability and also that this ability shows variability between individuals, given that variation is a driver of evolution. Men and women were assumed (no measurements or tests were involved) to exhibit the same bell-shaped ("normal") distribution of intelligence, an assumption Galton justified by analogy with height. Using the same analogy, he assumed the female distribution to have a lower average. Consequently, in *Hereditary Genius*, women feature solely as the mothers or wives of male geniuses.

Galton maintained the view that natural ability is a combination of intellect, personality, and moral traits, such as capacity, zeal, and the power to do laborious work. With respect to morals, he wrote that it is the nature of all of us to believe blindly in what we love, rather than in what we think most wise:¹⁸

We are indignant when others pry into our idols, and criticize them with impunity, just as a savage flies to arms when a missionary picks his fetish to pieces. Women are far more strongly influenced by these feelings than men; they are blinder partisans and more servile followers of custom.

The invention of a single, general form of intelligence, or natural ability, allowed Galton and his followers to compare men and women on a single dimension, similar to how he compared humans of different racial categories and even animal species. For instance, he conjectured that what he

¹⁶ Daston (1992).

¹⁷ Galton (1869/1979).

¹⁸ *Ibid.*, p. 196.

called the “negro race” differed from the “Anglo-Saxons” in their lower mean, not in the nature of their intelligence, and that certain gifted dogs had superior intelligence to some human “idiots and imbeciles.”¹⁹

Today, the idea of a general kind of intelligence is mostly related to Charles Spearman’s “g” factor.²⁰ In fact, Spearman was strongly influenced by Galton, and his main statistical tool was correlation, developed by Galton. Like Galton, he thought that high-sensory discrimination and high intelligence are part of the same universal intellectual function. Unlike Galton, however, Spearman steered clear of prejudices about women or non-whites being genetically inferior in their intelligence.

The Failure to Measure Intelligence

After Galton had invented the concept of general intelligence, he tried to measure it in his Anthropometric Laboratory in London, which opened in 1884. He started with the hypothesis that intelligence, being inherited, can be found in the mind and body – in the entire nervous system. From that perspective, greater sensory acuity would be the external sign of higher intelligence. Inspired by Galton, James McKeen Cattell established another anthropometric laboratory at Cambridge University, which also focused on sensory acuity. However, a student of Cattell’s, Clark Wissler, could not find a clear relationship between sensory acuity and mental ability when looking at college freshmen’s grades.²¹ Moreover, the various acuity measures did not appear to correlate with each other.²² Rather than acknowledging this failure as an invalidation of his hereditary theory of intelligence, Galton assumed a need for better measures of innate ability. His quest failed.

The key to measuring intelligence was found later in the work of Alfred Binet and Théodore Simon in France. In contrast to Galton and his followers, however, neither Binet nor Simon conceived intelligence as fixed or inherited, and Simon protested against the misuse of their test in England and the USA for measuring an allegedly inherited ability.²³

How Women’s and Men’s Average Intelligence Were Made Equal

Binet, a member – and, later, director – of the Free Society for the Psychological Study of Children, was concerned about the unreliable

¹⁹ Ibid., p. 338 (on race); p. 36 (on “idiots”).

²⁰ Spearman (1904).

²¹ Wissler (1901).

²² See Blum (1978); Sternberg (1990).

²³ Wolf (1973).

diagnoses of children with intellectual disabilities in France. One and the same child might be classified according to the categories back then as “imbecile,” “idiot,” “feeble-minded,” or “degenerate” in different certificates.²⁴ Around 1899, Binet set out to classify these children in an objective way with scientific precision. His goal was to place children with intellectual disabilities in special schools geared to improve their abilities, as in the German school system at the time, and also to ensure that children without any intellectual disabilities would not be placed in special classrooms solely because they were behaviorally challenging. But Binet had no coherent idea how to measure intelligence. Like Galton, he searched in vain for correlations with sensory acuity and tried almost everything else that seemed viable, including assessing intelligence on the basis of facial features (physiognomy), measurements of the head (cephalometry), and handwriting (graphology). For instance, he presented handwriting samples from convicted murderers mixed with those from normal citizens and asked expert graphologists for character assessments, only to find out that even the most eminent experts arrived at disastrously false assessments.²⁵ The results were consistently disappointing. It remained a mystery what intelligence was and how to measure it.

Eventually, however, Binet and Simon found an ingenious answer to the question of finding a test that correlated with teachers’ assessments. They developed questions about subjects that mirrored what was taught at school, such as reasoning skills, knowledge, memory, and attention. Children’s answers to these questions now correlated with their school grades as well as with teachers’ evaluations. By 1905, Binet and Simon had their first test of intelligence for classifying intellectually challenged children into several levels of developmental delay; in 1908, the test was revised and called a test of the “development of intelligence among children.” Note that the test was intended to sort children into categories, not to assign them a single number such as an intelligence quotient. It was also not intended to measure innate intelligence, but to replace teachers’ and physicians’ unreliable diagnoses of children with intellectual disabilities, as a “means of prophylaxis, a means of escaping conscious and unconscious error.”²⁶

Binet and Simon’s test questions still reflected the meaning of intelligence as a combination of intellect, character, and moral traits. For instance, the test included questions such as: “If you are late for school, what would you do?” and “Why should one judge a person by his acts

²⁴ Binet & Simon (1905/1973).

²⁵ Wolf (1973).

²⁶ Binet & Simon (1907/1914), p. 10.

rather than by his words?" Today, one might call this social intelligence, but Binet and Simon thought of social judgment as inseparable from intelligence. Now they had a test, but without a theory of intelligence, apart from a loose definition of intelligence as "judgment, otherwise called good sense, practical sense, initiative, the faculty of adapting oneself to circumstances. To judge well, to comprehend well, to reason well, these are the essential activities of intelligence."²⁷ Shortly before his death, Binet (1911) wrote: "Thus we return to our favorite theory: intelligence is marked by the best possible adaptation of the individual to his environment" and "to this we really do not want to add another thing."²⁸ To which his biographer Theta H. Wolf added: "How strikingly inept is such a pronouncement if we think of the excellent 'adaptation' to their environment of mice and moose!"²⁹ Measuring without precisely knowing what one is measuring has been, and still is, one of the striking features of research on intelligence. This feature conveniently allowed researchers to adjust the facts about female intelligence.

*Binet's Intelligence Test Crosses the Atlantic and Becomes Seen As a
Test of Genetic Ability*

After obtaining his PhD under G. Stanley Hall at Clark University, Lewis Terman joined the faculty at Stanford University and gained a reputation as *the* leading US researcher on intelligence. Terman was more interested in gifted children than in intellectually challenged ones. In line with Hall and Galton, he firmly believed that intelligence was inherited. He translated Binet and Simon's test into English, added and deleted some questions, and published the product in 1916, which became known as the *Stanford-Binet Intelligence Scales*.

Yet, Terman had made important alterations that went largely unnoticed in the USA, but were to have damaging implications. He named the test an *IQ test* (the term was originally introduced by the German psychologist William Stern), where IQ was the ratio between mental age and chronological age. He believed that whatever the test measured was fixed and inherited, or at least predominantly so. Whereas Binet and Simon thought of the test as a means to send children with intellectual disabilities to special schools so that they could ideally be channeled back into normal classrooms, Terman instead advocated for special institutions and the

²⁷ Binet & Simon (1905/1973), pp. 42–43.

²⁸ Binet (1911), p. 172.

²⁹ Wolf (1973), p. 210.

sterilization of the “mentally retarded.”³⁰ Terman had a strongly biased vision of what would happen once his test was widely applied: “There will be discovered enormously significant racial differences in general intelligence, differences which cannot be wiped out by any scheme of mental culture.”³¹

Under the leadership of Robert Yerkes, president of the American Psychological Association and a member of the Eugenics Record Office’s Committee on the Inheritance of Mental Traits, the *Army Alpha and Beta Tests*, based on Terman’s IQ test, were applied to 1.75 million men in World War I.³² Yerkes and his staff were convinced that the test measured native intelligence, even though it included items such as “The Overland car is made in Buffalo/Detroit/Flint/Toledo.”³³ They recommended about 8,900 men with low test results to be immediately discharged from service, many of whom were foreign born or illiterate. The Army officers disagreed with the psychologists, pointing out that these men would become good soldiers after training.³⁴ Nevertheless, Yerkes hailed the test a great success, despite little evidence that it had made recruiting more efficient or had contributed to winning the war. In fact, it was the war that helped to win publicity for mass testing – if only because the psychologists had shown that such testing could be accomplished. On that wobbly basis, IQ testing spread across the USA.

Binet did not live to see what happened to the Binet-Simon test once it crossed the Atlantic, but Simon did. He objected to the term *IQ* because it suggested a fixed, inherited mental age. In interviews with Binet’s biographer Theta Wolf, Simon even called the term and its genetic interpretation a betrayal (“trahison”) of their test’s original objective.³⁵

Men and Women Are Assigned the Same Mean Intelligence

Without much fanfare, Terman eradicated the idea that females have lower average intelligence. In his revised Stanford-Binet test, he deleted questions for which boys and girls had different success rates and balanced the rest so that, on average, girls ended up with the same IQ as boys. Terman was not particularly explicit about this correction, nor about its reasons. But his decision finally made women equal to men in terms of IQ, at least on average.

³⁰ Minton (1988), p. 149.

³¹ Terman (1916), p. 92.

³² Carson (2007).

³³ Minton (1988), p. 70.

³⁴ Ibid., p. 73.

³⁵ Wolf (1973), p. 203.

What was Terman's motivation? Terman and Maude A. Merrill later explained that they plotted the difficulties of each item against age groups "for the sexes separately as a basis for eliminating tests which were relatively less 'fair' to one sex than the other."³⁶ Moreover, "a considerable number of those retained show statistically significant differences in the percentages of success for boys and girls, but as the scales are constructed these differences largely cancel out."³⁷ The explanation of "fairness" appears strange in the face of Terman's intention to measure largely genetic differences in intelligence. And fair to whom? Were boys or girls originally better, and whose mean was upgraded? Terman and Merrill did not say.

Others proposed that Terman made the means equal to reckon with the fact that girls usually perform better in school or in response to pressure generated by the increasing women's movement of the period.³⁸ A third explanation is that Terman, working closely with a large number of women coworkers (according to his biographer, Henry Minton, sometimes too closely), was influenced by them. Yet, all three explanations assume that boys initially tested better than girls and that item deletion served to upgrade the girls' average. Who really did perform better in the original set of test, girls or boys?

It took me a while to find an answer in Terman's writings. It appeared years later, in a different context, in the study on gifted children by Terman and Melita Oden, hidden in a side remark on another topic, the question of why there were more boys than girls in the group of gifted children. Terman and Oden discussed the possibility of a nomination bias (teachers nominate more boys than equally gifted girls) and also the possibility of "a real average superiority of boys in the intellectual function tested."³⁹ They concluded that such a real average superiority is unlikely because for the 905 subjects on whom the 1916 Stanford-Binet test was standardized, the mean IQ was slightly higher in girls. In other words, Terman appears to have found that girls had higher average scores in his intelligence test than boys and then deleted items and balanced others to lower the mean of the girls to match the inferior mean of the boys!

One might ask what would have happened if girls had had the lower scores. Would Terman also have deleted items to even the averages out? If not, the test might have been standardized such that females' average IQ was a few points lower than that of males.

³⁶ Terman & Merrill (1937), p. 22.

³⁷ *Ibid.*, p. 34.

³⁸ Blum (1978).

³⁹ Terman & Oden (1947), p. 13.

Terman's decision to make the average IQ of males and females equal put an end to the second idea of a peculiarly female intelligence. It also illustrates the deep problem of how to measure something in the absence of a theory, where there is wiggle room to make decisions about test items that produce the result one favors – for fairness or whatever other reasons. In principle, Terman could have designed a test in which women are superior to men or where certain cultures or races are superior or inferior to white Americans. The problem is this: One can measure whether women and men differ in a specific and clearly defined task, such as memory span. But if one has neither a clearly defined task nor a theory and, instead, selects dozens of test items and adds the points up to determine an IQ, there are many degrees of freedom that allow for tinkering with the test to fit its result with preconceived beliefs and biases.

This key problem of measuring IQ is not always acknowledged. Consider Hans-Jürgen Eysenck, who once was the most frequently cited living psychologist and one of the most controversial intelligence researchers. In his *The Intelligence Controversy* with Leon Kamin, he reified the equal averages, complaining that psychologists “are said to have selected items in such a way that equal scores are achieved regardless of whether there might or might not be genuine differences between the sexes. This accusation is false.”⁴⁰ He continued: “Given that unselected items give the sexes equal IQ scores, it was only reasonable for other test designers to avoid bias in favour of one or the other sex.” However, there is no such thing as “unselected” items in the absence of a theory of what intelligence is and how it can be measured. Terman himself occasionally reified the equality of mean IQ to support women's equality. In *Sex and Personality*, Terman and Catherine Cox Miles wrote: “Intelligence tests, for example, have demonstrated for all time the falsity of the once widespread prevalent belief that women as a class are appreciably or at all inferior to men in the major aspects of intellect.”⁴¹ All in all, Terman's IQ test ended the view that females have lower average intelligence than males so that men and women were finally seen as equally intelligent – at the expense of favoring racial prejudice.

How Differences in Intelligence Become Differences in Personality

In the introduction to *Sex and Personality*, Terman and Miles noted that it appears impossible to explain sex differences in behavior wholly in terms of

⁴⁰ Eysenck & Kamin (1981), pp. 40–41.

⁴¹ Terman & Miles (1936), p. 1.

biological factors and complained that the concepts of masculinity and femininity are even more vague than the 19th-century concepts of intelligence.⁴² By way of example, they referred to the stereotype of the “occidental” woman whose moral life is shaped less by principles than by personal relationships and whose everyday behavior is more determined by emotion, submissiveness, and inferior steadfastness of purpose.

Nevertheless, Terman and Miles did not present a theory that replaced the vagueness and stereotypes to which they objected. How then could they measure personality differences between men and women? Terman and Miles came up with an ingenious solution, which was initiated as subtly as Terman’s strategy to discard test questions had been. It turns out that the discarded questions ended up in their “masculinity-femininity scale.” That action guaranteed differences between males and females on the new scale, which contained, among others, questions on interests such as movies and amusement, opinions such as “The unmarried mother deserves the scorn she gets” and “Blondes are less trustworthy than brunettes,” and information such as “The most gold is produced in Alaska/NY/Tennessee/Texas.” Once seen as items that measured inherited intelligence, these now served to measure personality and gender-specific knowledge. In the absence of a theory of intelligence that determines what questions are relevant, one-and-the-same item can be applied to measure sex differences in intelligence or in personality. In various forms, the masculinity-femininity scale remains in use and is still presented as measuring sex differences in personality.

Larger Variability in IQ Justifies Male Superiority

In 2006, Harvard president Larry Summers resigned from his position in the wake of a no-confidence vote by his faculty. Among the reasons cited by the faculty was a remark he had made regarding women’s intelligence and ability. On the question of women’s aptitude for science, Summers said: “It does appear that on many, many different human attributes – height, weight, propensity for criminality, overall IQ, mathematical ability, scientific ability – there is relatively clear evidence that whatever the difference in means – which can be debated – there is a difference in the standard deviation, and variability of a male and a female population.”⁴³ From that he drew the conclusion that the greater variability of males

⁴² Ibid., pp. v–vi.

⁴³ Summers (2005).

explains why top universities such as Harvard hired relatively few women as professors.

Summers' statement simply repeated a hypothesis discussed in psychological research for over a century: that the variability of women's physical and mental traits, including IQ, is smaller than that of men. This variability hypothesis both explains and justifies observations that there are more male geniuses than female ones and also explains why there are more male idiots at the other end of the IQ distribution.

After Galton replaced the first version of intelligence – that men's and women's mental abilities were at opposite poles, such as reason versus intuition – with one common intelligence, and Terman, in turn, put an end to the subsequent idea of average differences, the only possible remaining difference on the bell curve was the variability, or standard deviation, in IQ. After all, a bell curve has only two parameters, mean and standard deviation. The variability hypothesis became the third and last bastion for the idea of a specifically female intelligence, contributing to Summers' fall. Its origins seem to lie in an observation by Darwin in the second edition of *Animal and Plants Under Domestication* that male animals tend to be more variable than females, although Darwin himself devoted little attention to this issue.⁴⁴ Instead, the claim of greater male variability was promoted by the English sexologist Havelock Ellis.

The Variability Hypothesis

Ellis rebelled against the conspiracy of silence surrounding the sexes and decided to devote his life to their scientific study. For him, women and men were different, but complementary – in contrast to Galton, who did not see much usefulness in women's lower average natural ability. In the first edition of *Man and Woman*, Ellis wrote: "From an organic standpoint, therefore, men represent the more variable and the more progressive element, women the more stable and conservative element, in evolution. It is a metaphorical as well as a literal truth that the center of gravity is lower in women and less easily disturbed."⁴⁵ (In the fourth and fifth edition, Ellis left out the "progressive element," indicating second thoughts about the generalizability of biological variation, particularly to politics.) He wrote that women's smaller stature approximated that of humans' ancestors, and that women – as in witches and soothsayers – preserved ancient custom and methods of intuitive knowledge. Women had "an

⁴⁴ Darwin (1893), p. 457. ⁴⁵ Ellis (1894), p. 367.

organic tendency to stability and conservatism, involving a diminished individualism and variability."⁴⁶ To exemplify, he made the case that women had opposed the French Revolution, albeit also noting that the revolutionary movement of Christianity was, to a considerable extent, furthered by women. While acknowledging that the facts are very complex and that the claim of absolute inferiority for either sex is untenable, Ellis nonetheless concluded: "It is undeniably true that the greater variational tendency of the male is a psychic as well as a physical fact."⁴⁷

Man and Woman received scant attention when it first appeared.⁴⁸ Yet, this changed when the statistician Karl Pearson vigorously attacked Ellis' variability hypothesis.⁴⁹ Pearson was a committed socialist and promoted feminism and eugenics, both of which were considered progressive and revolutionary at the time. Pearson argued that the claim of greater male variability contradicts Darwin's theory of evolution by natural selection, which emphasizes variability as one of the driving forces of evolution, but postulates that the more intense the struggle is, the less is the variability. Therefore, he expected men, not women, to be less variable. Next, he criticized Ellis' inconclusive evidence, based almost entirely on pathological variation such as criminality and color-blindness. Finally, Pearson contended that measuring the variability of absolute variables such as the length of bones (as opposed to ratios such as cephalic index) by the standard deviation, as Ellis did, was an error. Instead, one needed to calculate the coefficient of variation, that is, the standard deviation divided by the mean. After all, women's bodies were smaller than men's and so, therefore, was the standard deviation of bodily measures. Pearson concluded from his own physical measures that the coefficient of variation is slightly larger for women, not smaller, reflecting their "slightly less severe struggle for existence."⁵⁰

In an Appendix in *Man and Woman*, Ellis rejected Pearson's "hostile" criticism at length, which Pearson did not deem worthy of a response. Pearson's sole reaction was a footnote in an article unrelated to variability, in which he noted that Ellis' response required no reply as Ellis did not appear to understand that scientific evidence, not vague generalities, was what counted.⁵¹ Afterwards, Pearson did not pursue the variability hypothesis any further.

Why did this bitter controversy over females' allegedly lower variability erupt? According to Ellis' biographer Phyllis Grosskurth, one likely reason

⁴⁶ Ibid., p. 369.

⁴⁷ Ibid., p. 370.

⁴⁸ Grosskurth (1980), p. 170.

⁴⁹ Pearson (1897).

⁵⁰ Ibid., p. 297.

⁵¹ Pearson & Lee (1903), p. 372.

was personal resentment. Many women of the time found Ellis, who with his flowing beard resembled “a combination of archetypal Father and sensual Faun,” irresistibly attractive.⁵² The South African writer Olive Schneider was one of the women upon whom Ellis had a strong influence, before she fell in love with Karl Pearson. Whatever the motivation, Pearson’s critique of the variability hypothesis in fact contributed to making the hypothesis popular.

Ever since, psychologists, biologists, and statisticians have debated the variability hypothesis. Whereas Ellis and Pearson related it to both physical and mental traits, psychologists have focused largely on intelligence. Quinn McNemar and Terman reported greater variability in boys on the Stanford-Binet and other tests, but, given the inconsistent evidence, were careful not to draw any general conclusions.⁵³ In 1932, Scotland undertook the ambitious project of testing all 11-year-old Scottish children with the goal of discovering the amount of mental deficiency in the country.⁵⁴ Because suppliers demanded too much money for the nearly 100,000 commercial tests, the Council used the *Morey House Test* in place of the Stanford-Binet test. The conclusion was that boys and girls did not differ in average IQ, but that the standard deviation of boys was one IQ point higher than that of girls. In 1947, the same project was repeated with all 11-year-olds at that time, and again the standard deviation was one point higher for boys. This appeared to support the hypothesis of both higher and lower male intelligence. Although this result was hailed as the most comprehensive demonstration of the greater variability of mental ability among males, the small difference in variability in the 1947 study was, in fact, mainly due to an excess of males with very low scores, not to male genius.⁵⁵ The primary impetus of the 1947 study (and that of 1932) was not variability, but rather the concern that the nation’s intelligence would decline because people with lower mental ability tended to have more children. Yet, the children scored no worse than those studied 15 years earlier; in fact, their average IQ went up by about one point in boys and three points in girls.

Follow-ups of the Scottish children study have shown similar, inconclusive results. In 1939, the Council found no significant difference in variability between boys and girls; in 1949, it reported slightly larger standard deviations in boys; and in 1958, it reported a greater proportion

⁵² Grosskurth (1980), p. xvi. ⁵³ McNemar & Terman (1936).

⁵⁴ Scottish Council for Research in Education (1933, 1939, 1949, 1958).

⁵⁵ Deary et al. (2009), pp. 21, 184–185.

of females than males at the lower end of the IQ scale. Thus, one could find support for or against the variability hypothesis in intelligence, depending on the age group and study. More fundamentally, findings about variability – like mean differences – always depend on how the test items are selected and weighted. Just as Terman made the means between males and females equal, one can select items to make the variability equal.

Outspoken advocates have presented greater male variability as a biological fact, possibly due to sex linkage, speculating that intelligence might be located on the X chromosome. According to this line of reasoning, intelligence in males can express itself without interference of a second X chromosome, thereby causing greater variability in IQ.⁵⁶ This ignores the fact that the same hypothesis could likewise be used to predict that females have higher average intelligence than men, thanks to their two X chromosomes, once again illustrating the utter arbitrariness of genetic explanations in the absence of a theory. Whereas the first two ideas about a peculiarly female intelligence had been conceived and debated virtually entirely by men, the variability hypothesis was challenged by an early generation of women scientists.⁵⁷ Helen Bradford Thompson conducted her own studies and criticized Ellis' conclusions.⁵⁸ Her critique of the variability hypothesis was widely read, yet had no equivalent impact. In the most systematic critique of the variability hypothesis at the time, Leta Stetter Hollingworth reported no evidence of this in her review of the literature.⁵⁹ Beth Wellman found in her review slight support for greater variability in boys, which, however, depended on the measure of variability used, the selection of children, and other details.⁶⁰ The variability hypothesis remains a matter of discussion. In her 2012 review of the state of art in sex differences in cognitive abilities, the former president of the American Psychological Association, Diane Halpern, accepted it, concluding "that females and males are very similar when we consider the average performance, and they are highly dissimilar when we consider performance at the high and low extremes."⁶¹

As with the question of whether males and females differ in their average IQ, the absence of a theoretical understanding of what a test actually measures opens the door to including or excluding items that make the mean and variance of IQ equal or different.

⁵⁶ Johnson et al. (2009); Lehrke (1978). ⁵⁷ Shields (1982). ⁵⁸ Thompson (1903).

⁵⁹ Hollingworth (1914). ⁶⁰ Wellman (1933). ⁶¹ Halpern (2012), p. 103.

Lessons Learned

In sum, the idea of a peculiarly female intelligence emerged in three different and unrelated versions: male–female polarities, female lower mean intelligence, and female lower variability. The idea that men and women occupy opposite poles on a continuum, such as analytic versus intuitive, is the oldest; it reigned for millennia. It began to fade away when Francis Galton invented intelligence (natural ability) as a single dimension, which later morphed into IQ or *g* (general intelligence), so that the minds of men and women now had the same quality, but with women having less of it. The idea that women have lower intelligence expired in the hands of Lewis Terman, who eliminated test items so that both males and females had the same average IQ – otherwise, female means would, in fact, have been higher. The third idea was that while the means are the same, woman’s variability is smaller, resulting in more male geniuses and idiots. This variability hypothesis is still debated today. It remains the last bastion of those who cling to the idea of male supremacy.

Despite the differences in these three ideas about a peculiarly female intelligence, their justifications are strikingly similar, and the supposed nature of women features prominently in all three. A woman’s mind was said to be determined by her reproductive biology, her body, her genes, and her naturally ordained functions. The first president of the American Psychological Association, G. Stanley Hall, staunchly believed that the female mind was created for nursing and motherhood, serving the production of men of genius and of daughters to bear future male geniuses.⁶² Education, he felt, would damage women’s reproductive organs, particularly coeducation in competition with men. Like many others at the time, Hall did not think of women as generally inferior, but instead idealized them. In his view, women who entered the men’s world of education and business became innocent victims of man’s evil nature, losing their purity and sainthood.⁶³

The historian of psychology, Edwin Boring, famously said that intelligence is whatever the IQ tests measure. But that is precisely the problem. The idea of a peculiarly female intelligence is a striking case of measurement without understanding what one is measuring, paired with the hope that sophisticated correlation statistics and factor analyses could fill this theoretical void. From Galton to Binet to Terman, researchers variously believed that one could measure intelligence in terms of sensory acuity,

⁶² Diehl (1986). ⁶³ Schofer (1976).

head size, facial features, handwriting, memory capacity, or knowledge of facts, or by asking questions about proper social behavior.

This absence of theory left too many points of entry for biases and preset convictions, to the detriment of many. Galton's vision was to promote the eugenics program: to detect the less well-endowed and prevent them from reproducing. Both Ellis and Pearson were early feminists but also proponents of eugenics, both of which were considered progressive movements at the time. Binet and Simon intended to give children with intellectual disabilities a second chance through special education. Yet, when adapted "to American conditions and needs," as the editor's introduction to the 1916 edition of Terman's *The Measurement of Intelligence* put it, their test came to serve the various goals of eugenics, sterilization, racism, feminism, and, last but not least, a multibillion testing industry.

Why Is History Relevant?

Knowing one's history provides an opportunity to learn from errors and to avoid repeating these. Differences between men and women, as well as their causes, have been an emotionally and politically charged topic for centuries. Firm convictions continue to be enforced in the guise of new technology. For instance, Diane Halpern warns that modern neuroscience is being misused to justify sex role stereotypes in how men and women think, a program dubbed "neurosexism."⁶⁴ Basing conclusions about human thinking and behavior on the firing of neurons or changes in blood oxygen levels entails a long leap in logic. Such leaps are not new; we have already seen one, for instance, in the argument that the smaller brain of females is responsible for woman's alleged intellectual inferiority. The stereotypes of the past also tenaciously survive in popular psychology bestsellers that present men and women as if they were alien species, as in *Men Are from Mars, Women Are from Venus*.⁶⁵ In a throwback to the view of women being submissive by nature, such books imply that a wife's role is to hide her intelligence, to admire and appreciate her husband, and to not offer him advice unless he asks. Online communities such as *Men Going Their Own Way (MGTOW)* and *TradWives* revive the traditional view that the position of men is above women.

What is the current consensus about differences in men's and women's cognitive abilities? According to Halpern, the list of differences is relatively small, and the similarities between the sexes are larger in number. Few of

⁶⁴ Halpern (2012), p. xi. ⁶⁵ Gray (1992).

the differences that have been claimed over the years are stable across age, task, and culture. Among the few exceptions are that women have better memories than men and excel in reading and verbal abilities, while males excel in science and mathematics.⁶⁶ What causes these differences is far from being understood.

This history of the idea of a peculiarly female intelligence can teach us several general lessons. The first is to beware of research that evaluates the sexes in terms of polarities and, in general, uses polarities as a means to understand the human mind. Second, beware of composite index numbers, such as IQ. Unless there is a strong theory, test items can be selected to verify any existing bias “scientifically.” Third, keep in mind that intelligence is about cognitive processes. All in all, we would be well advised to replace polarities and IQ numbers with the study of the actual processes underlying intelligent behavior, a scientific research agenda that would also leave little room for individual and cultural biases.

Beyond Polarities and IQ: Intelligent Decision Processes

The history of the idea of a peculiarly female intelligence shows, in my view, that the field of sex differences in intelligence, and of intelligence in general, could benefit from a fresh start. Herbert Simon’s and Alan Newell’s work on heuristics and artificial intelligence (AI), which has inspired my own research on both intuitive and deliberative heuristic decision-making, offers such a new framework. Heuristics are strategies that help agents to make decisions and solve problems in an intelligent and efficient way.⁶⁷ After all, what we call intelligence manifests itself in the quality of the decisions we make. In Part II, I describe the research agenda in more detail, which centers on two questions: (i) What is the repertoire of intelligent strategies (such as heuristics) at a person’s disposal for making decisions? And (ii) what is a person’s ability to choose a proper strategy for the situation at hand? In this framework, intelligence has a very concrete meaning that connects cognitive abilities with behavioral strategies, namely, the “adaptive toolbox” of strategies available and the ability to choose a strategy wisely to achieve a particular goal.

Although the study of intelligent heuristics is well established, it has had a blind spot for sex differences in how males and females search for information, when they stop searching, and how they make or delay

⁶⁶ Halpern (2012); Halpern & Wai (2020), pp. 119, 126–127.

⁶⁷ Gigerenzer & Gaissmaier (2011); Gigerenzer et al. (2011).

decisions. One exception is the work of Joan Meyers-Levy and Barbara Loken, who reported that, in consumer choice, females search more extensively for information than males, while males are more selective in search and rely on faster stopping rules.⁶⁸ Moreover, they concluded that females are more sensitive to environmental cues, whereas men more often ignore these and rely on the same heuristics across contexts, indicating less ability in adaptive choice. As for social heuristics, they found that women are more likely to base decisions on trust and are more likely to be trusted. Note that these are preliminary findings, but they indicate a different kind of question to pursue: abandon studying polarities and differences in IQ test outcomes and, instead, ask whether there are concrete differences in the way males and females search for information and make decisions.

The Myth of Female Intuition and Male Reason

It is telling that quite a few smart men in the social, biological, and medical sciences insisted on female inferiority as a scientific fact, in terms of both intellect and morals. Many of the tribulations mankind has had to endure were attributed to Eve's weakness when facing the serpent. After females were allowed to attend schools and universities during the 20th century, it became evident that science had previously been informed by prejudice. The opposition between female intuition and male reason then disappeared from the social sciences and women were no longer accused of naivete, hasty conclusions, lack of logic, or low moral principles. Nevertheless, traces of these prejudices still surface occasionally. Why are so few females hired as professors in philosophy departments? Some male philosophers have argued that intuition is gendered, and that women have wrong philosophical intuitions more often than men.⁶⁹

In the course of time, women and men have become increasingly equal partners – but not intuition and rationality.

⁶⁸ Meyers-Levy & Loken (2015). ⁶⁹ Tripodi (2015).