BI-11 Within-sex variation and individuals

The systematic study of human sex differences is frequently hindered by a conceptual failure to distinguish between statistical group averages and individual realities. Group-level averages—which form the basis of most scientific publications and public discourse—frequently obscure the profound within-sex variation that characterises almost every physiological, behavioural, and cognitive trait in the human species. To navigate this complexity without falling into either false dichotomies or false equivalencies, an analytical framework must separate biological and psychological phenomena into three distinct layers: physiology (the measurable, typically highly dimorphic foundation), behaviour (the statistical expressions resulting in overlapping distributions), and culture (the variable social roles and environments that modulate expression).

Understanding how men and women differ, and critically, how they do not, requires moving beyond binary categorisations to examine the exact statistical distributions, the degree of overlap, and the properties of the distribution tails. This report systematically quantifies within-sex versus between-sex variation across hormones, neuroanatomy, and behaviour, presenting the strongest evidence from peer-reviewed primary literature, systematic reviews, and meta-analyses.

The Statistical Tools for Communicating Differences Honestly

To distinguish what is universally true of an entire sex from what is true only on average, a precise statistical vocabulary is required. Standardised mean differences, most commonly expressed as Cohen's d or Hedges' g, quantify the distance between two group means in units of pooled standard deviation[1]. A positive d value typically denotes a male advantage or higher male average, while a negative d indicates a female advantage. However, raw effect sizes are abstract and frequently misinterpreted by both researchers and the public. Translating these metrics into concrete distributional realities requires specific mathematical tools.

The Probability of Superiority (PS), also known as the Common Language Effect Size (CLES), was formally proposed by McGraw and Wong in 1992 in the *Psychological Bulletin*[1]. The CLES translates Cohen's d into a practical probability. It answers a highly specific question: If one male and one female are drawn at random from the population, what is the probability that the individual from the group with the higher mean will possess a higher value on the trait than the individual from the lower-scoring group[1]? Under the assumption of normal distributions with equal variances, the CLES is calculated using the standard normal cumulative distribution function. For example, a medium effect size of $d = 0.50$ translates to a 64% chance that a

randomly selected member of the higher-scoring group will outscore a randomly selected member of the lower-scoring group[1].

The Overlap Coefficient (OVL) quantifies the total area of intersection between two probability distributions. An OVL of 1.0 indicates perfect overlap and identically matched distributions, while an OVL approaching 0 indicates complete separation[1]. Measuring distributional overlap is critical because it forces analysts to recognize similarity alongside difference.

Finally, Cohen's U_3 statistic calculates the proportion of the lower-scoring group that is exceeded by the average (median) member of the higher-scoring group[6]. This provides a direct answer to questions regarding what proportion of one sex falls outside the typical range of the other.

To illustrate how these tools communicate differences honestly, the standard benchmarks for Cohen's d can be mapped onto their exact distributional realities, assuming normality and equal variances.

Standardised Difference (Cohen's d)	**Conventional Effect Size Label (Cohen, 1988)**	**Probability of Superiority (CLES)**	**Overlap Coefficient (OVL)**	**Proportion of Lower Group Exceeded by Higher Group Mean (U3)**
0.20	Small	56%	92%	58%
0.50	Medium	64%	80%	69%
0.80	Large	71%	69%	79%
1.00	Very Large	76%	62%	84%
2.61	Extreme (e.g., Human Height)	96%	19%	99.5%

While Cohen's d is sufficient for isolated, univariate traits, psychological and physiological profiles are fundamentally composite. To measure global, multidimensional differences, analysts utilise Mahalanobis D. This multivariate effect size aggregates univariate differences while strictly accounting for the covariance (correlations) between the traits[2]. Methodological work has repeatedly demonstrated that two groups may exhibit massive overlap on ten individual traits (resulting in a small Cohen's d for each), but when those traits are combined

into an interconnected profile, the groups may separate almost entirely, resulting in a large Mahalanobis D and very low multivariate overlap[7].

Finally, comparing the dispersion or spread within the sexes requires the Variance Ratio (VR). The VR is calculated by dividing the male variance by the female variance. A VR of 1.0 indicates equal variability. Because the extreme tails of a normal distribution are highly mathematically sensitive to variance, even a seemingly negligible increase in variance (e.g., a VR of 1.1) results in a massive overrepresentation of that specific group at the extreme upper and lower percentiles of the overall distribution[9].

Layer I: Physiology (Measurable)

The physiological layer encompasses genetics, gonadal development, endocrinology, and neuroanatomy. It is at this measurable foundation that human beings exhibit the most categorical dimorphism, yet a nuanced analysis reveals that substantial within-sex variation still dictates individual biological reality.

Intersex Variation and the 3G-Sex Concept

Categorical biological sex is determined by the alignment of genetic (chromosomal), gonadal, and genital features—a biological triad sometimes referred to in the literature as 3G-sex[12]. In the vast majority of the human population, these three levels align consistently as either unambiguously female (XX chromosomes, ovaries, typical female genitalia) or unambiguously male (XY chromosomes, testes, typical male genitalia)[12].

The prevalence of intersex conditions—variations of sex development where this triad does not neatly align—has been the subject of intense academic and cultural controversy. In a highly influential paper published in the *American Journal of Human Biology* in 2000, Blackless, Fausto-Sterling, and colleagues reviewed medical literature spanning from 1955 to 2000. They concluded that the frequency of individuals deviating from the platonic ideal of male or female was approximately 1.7% of all live births[13]. This specific 1.7% figure—which makes intersex traits roughly as common globally as having red hair—has since been widely cited across sociology, gender studies, and popular media to argue that biological sex is a continuous spectrum rather than a binary[15].

However, the 1.7% estimate has faced severe methodological critique from medical researchers. In 2002, Leonard Sax published a direct rebuttal in the *Journal of Sex Research*, demonstrating that the Blackless and Fausto-Sterling figure relied on an exceptionally broad definition of "intersex" that included medical conditions where chromosomal sex is entirely consistent with phenotypic sex and where genitalia are unambiguously male or female[16]. Specifically, Sax pointed out that 1.5 percentage points of the 1.7% estimate (nearly 88% of the total cases) were composed of individuals with Late-Onset Congenital Adrenal Hyperplasia (LOCAH)[17]. LOCAH is a condition that does not cause ambiguous genitalia at birth and frequently remains entirely asymptomatic until puberty or adulthood[17].

Furthermore, Fausto-Sterling's estimate included sex chromosome aneuploidies such as Klinefelter syndrome (XXY, estimated at 0.09%) and Turner syndrome (X0, estimated at 0.03%), which produce unambiguously male and unambiguously female phenotypes, respectively[13].

When restricting the definition of intersex to conditions involving genuinely ambiguous genitalia or true chromosomal and phenotypic discordance (such as Complete Androgen Insensitivity Syndrome or Ovotestis), Sax calculated the true prevalence to be approximately 0.018%, or roughly 1 in 5,500 births[17].

This 0.018% figure is nearly 100 times lower than the Fausto-Sterling estimate[17]. Thus, while variations in sex development present vital, sensitive medical and ethical considerations requiring highly individualised care, the physiological baseline of the human species remains overwhelmingly dimorphic. Over 99.98% of individuals fall into unambiguous male or female developmental categories at birth[12].

Hormonal Distributions: Averages, Extremes, and Behavioural Prediction

Endocrinology displays both categorical between-sex differences and extensive within-sex variation. Testosterone, the primary androgen, is a fundamental driver of male physiological organization, but its presence and impact are heavily mediated by developmental timing, environmental context, and genetic receptor sensitivity.

Testosterone exposure operates in critical, time-delimited developmental windows. During gestation, fetal testosterone peaks between 11 and 14 weeks. This prenatal surge drives the masculinisation of the internal and external genitalia and initiates the early structural organization of the developing brain[19]. Indirect markers of this prenatal androgen exposure, such as the 2D:4D digit ratio (the length of the second digit divided by the fourth digit), show highly replicated sex differences. A meta-analysis by Hönekopp and Watson (2010) confirmed that males have a significantly lower 2D:4D ratio on average than females, reflecting higher prenatal testosterone exposure, with the effect being most pronounced in the right hand[21].

Following birth, the withdrawal of placental estrogens triggers a secondary, massive surge in gonadotropins (Luteinizing Hormone and Follicle-Stimulating Hormone) and testosterone in males, a phenomenon termed "minipuberty"[19]. Between 1 and 6 months of age, healthy male infants experience a dramatic testosterone spike, with circulating levels reaching between 100 and 400 ng/dL[22]. This range is remarkably high, comparable to early-to-mid pubertal levels[23]. This postnatal testosterone surge correlates physically with penile growth and Sertoli cell proliferation, and neurologically with the masculinisation of specific neurobehavioural pathways[19]. In stark contrast, female infants experience a distinctly different minipuberty marked by fluctuating surges in estradiol to stimulate follicular maturation, but they maintain negligible testosterone levels throughout infancy[23].

In adulthood, the between-sex variance in testosterone is absolute. Male baseline testosterone is typically 10 to 15 times higher than female baseline testosterone. Because these distributions are entirely disjoint, the overlap coefficient is virtually zero. No healthy adult female possesses a baseline testosterone level that falls within the normal healthy range of an adult male. However, the variance *within* the male population is vast. Adult men vary among themselves from roughly 300 ng/dL to over 1,000 ng/dL[19].

A central question in behavioural endocrinology is the extent to which this massive within-sex variation in testosterone predicts individual behaviour, specifically regarding aggression,

risk-taking, and status-seeking. The "Challenge Hypothesis," originally developed in avian and non-human animal models by Wingfield et al. (1990), posits that testosterone rises in response to status threats and facilitates aggressive, competitive behaviour to secure reproductive advantages[26].

Translating the Challenge Hypothesis to human behaviour has yielded remarkably weak associations. In a widely cited systematic review and meta-analysis published in the *Review of General Psychology*, Archer, Graham, and Sommer (2005) revealed that the overall correlation between baseline testosterone and human aggression is extremely weak, yielding an effect size of $r = 0.08$ to 0.14 across studies[27]. This means that knowing a man's baseline testosterone level provides virtually no predictive power regarding how aggressive he will be in his daily life.

To explain these small effect sizes, researchers proposed the "Dual-Hormone Hypothesis" (Mehta & Josephs, 2010). This theory suggests that testosterone only predicts status-seeking and aggression when cortisol (the primary stress hormone regulated by the Hypothalamic-Pituitary-Adrenal axis) is low[26]. If an individual is highly stressed or anxious, elevated cortisol acts antagonistically to block the behavioural effects of testosterone[30].

Despite the theoretical elegance of the Dual-Hormone Hypothesis, its empirical robustness has failed to replicate consistently. A highly exhaustive meta-analysis by Dekkers et al. (2019, *Psychoneuroendocrinology*), comprising 33 studies and 8,538 participants, found only marginal, statistically borderline support for the interaction of testosterone and cortisol on status-relevant behaviour[26]. The interaction effect size was exceptionally small ($r = -0.061$, $p = 0.026$)[26]. The authors cautioned that this effect is too weak to be useful for predicting individual behaviour[26]. Furthermore, a subsequent *p*-curve analysis by Grebe et al. (2019, *Hormones and Behavior*) indicated that the published studies supporting the Dual-Hormone Hypothesis suffer from exceptionally low statistical power (estimated at just 16%), strongly suggesting publication bias and warning that the core findings have not robustly replicated[30].

Adding further complexity to the hormone-behaviour link is genetic variation in androgen receptor (AR) sensitivity. Circulating testosterone must bind to receptors to exert an effect. The AR gene, located on the X chromosome, contains a highly variable polyglutamine tract encoded by a repeating sequence of cytosine-adenine-guanine (CAG)[32]. A lower number of CAG repeats (a shorter sequence length) is strongly associated with increased receptor transcriptional activity and heightened cellular sensitivity to testosterone[32]. Consequently, an individual with moderate circulating testosterone but highly sensitive receptors (e.g., 16 CAG repeats) may exhibit significantly greater physiological and behavioural androgenisation than an individual with high circulating testosterone but highly resistant receptors (e.g., 28 CAG repeats)[34]. This biological reality guarantees that simple measurements of circulating hormone levels in the blood or saliva can only ever explain a minor fraction of the behavioural variance among men.

Finally, within-sex variation in testosterone is not static; it is deeply responsive to social context and life history transitions. In a landmark longitudinal study published in the *Proceedings of the*

National Academy of Sciences (2011), Gettler, McDade, Feranil, and Kuzawa provided definitive evidence that the transition to active fatherhood causes a severe, statistically significant decrease in circulating testosterone in human males[37]. Over a 4.5-year period in a cohort of men in the Philippines, those who became partnered fathers experienced much steeper declines in waking and evening testosterone compared to their peers who remained single non-fathers[38]. This replicated physiological downregulation highlights that human male endocrinology is highly plastic, adapted to suppress mating effort and aggression in order to promote long-term paternal investment and infant bonding[38].

The Neuroanatomy Debate: Dimorphism vs. The Brain Mosaic

The question of whether human brains are sexually dimorphic remains arguably the most fiercely debated topic in contemporary neuroscience. High-powered neuroanatomical studies consistently find average structural differences between the sexes, but the interpretation of these averages for individual brains—and whether a true "male brain" exists—is highly contested.

The largest single-sample study of structural and functional sex differences in the adult human brain was conducted by Ritchie et al. (2018, *Cerebral Cortex*) using high-resolution MRI data from 5,216 UK Biobank participants (2,750 females, 2,466 males; mean age 61.7 years)[41]. The study identified undeniable group-level average differences. Males possessed significantly larger raw total brain volumes, larger cortical surface areas, and higher white matter fractional anisotropy (indicating different microstructural organisation)[42]. Conversely, females exhibited significantly thicker cortices and higher white matter tract complexity[42].

Because male brains are, on average, 10% to 15% larger in total volume than female brains, researchers must adjust for this overall size disparity to isolate true structural dimorphism. Ritchie et al. found that even after strictly adjusting for total brain volume and height, highly specific regional sub-differences remained robust. For instance, females possessed relatively smaller adjusted volumes in the amygdala and pallidum compared to males[42]. Importantly, Ritchie et al. also identified greater male variance across almost all raw structural measures, providing strong neuroanatomical support for the Greater Male Variability Hypothesis[42].

Despite these highly reliable group-level differences, the fundamental debate centers on the individual: does a discrete "male brain" or a "female brain" actually exist within any single person? Daphna Joel and colleagues sought to answer this in a highly publicised 2015 study in the *Proceedings of the National Academy of Sciences*[46]. Joel et al. examined MRI datasets from over 1,400 individuals. They calculated the univariate distribution for various brain regions and arbitrarily defined "male-end" features as those falling within the extreme 33% most common in males, and "female-end" features as those in the extreme 33% most common in females[46]. They then calculated the *internal consistency* of individual brains—meaning the proportion of individuals who possessed *only* male-end features or *only* female-end features across their entire neural architecture[46]. Joel et al. found that perfect internal consistency was exceedingly rare. Depending on the exact dataset and threshold used, only 1% to 6% of human brains were internally consistent[46]. Instead, the vast majority of individuals possessed "substantially variable" brains containing a mixed mosaic of both male-end and female-end features[46]. The

authors boldly concluded that "human brains do not belong to one of two distinct categories: male brain/female brain," arguing that biological sex does not produce categorical dimorphism in the central nervous system[12].

The "mosaic brain" conclusion was immediately met with intense, mathematically rigorous pushback from leading methodologists and evolutionary psychologists. In 2016, Del Giudice et al. published a formal critique in the *Proceedings of the National Academy of Sciences*, arguing that Joel et al.'s methodology was mathematically guaranteed to obscure sexual dimorphism[48]. Del Giudice and colleagues demonstrated that assessing brain variables individually (univariate thresholding) completely ignores the vital covariance and structural correlations between brain regions[48].

To prove this, Del Giudice et al. generated simulated datasets of highly dimorphic animal populations (e.g., species of monkeys known for extreme, non-overlapping facial dimorphism). They proved that if random biological noise is added to the data, applying Joel's univariate mosaic methodology results in the exact same "lack of internal consistency," falsely concluding that the highly dimorphic monkeys possess "mosaic faces"[48].

By applying proper multivariate pattern recognition (machine learning) to the exact same neuroimaging data types, Del Giudice et al. (2016) and Chekroud et al. (2016, *Proceedings of the National Academy of Sciences*) demonstrated that algorithms could correctly classify an individual's biological sex from their brain scan with 69% to 93% accuracy[44]. Furthermore, Del Giudice calculated the multivariate Mahalanobis D for these brain structures to be approximately $D = 1.6$ to 1.7. This indicates significant global separation between the male and female multivariate distributions (roughly 24% overlap), definitively proving that brain structure is sexually dimorphic when viewed as a whole[48].

The strongest scientific evidence therefore demands a synthesis of these two perspectives. At the level of isolated, single brain regions viewed univariately, female and male brains overlap extensively, and a perfectly internally consistent "extreme male" brain is statistically rare. However, when the brain is properly analysed as an interconnected, multivariate system, male and female neural architectures separate reliably, allowing highly accurate classification and proving robust dimorphism[44].

Layer II: Behaviour (Statistical, Overlapping Distributions)

Moving from physiological architecture to psychological output, sex differences express themselves not as absolutes, but as highly statistical, overlapping probability distributions. In human behaviour, undeniable average differences exist, but the range of within-sex variation is usually far larger than the distance between the two group means.

Cognitive and Spatial Traits

Human cognitive abilities display largely equal averages between the sexes, particularly regarding the factor of general intelligence (g)[50]. However, specific, highly circumscribed

cognitive domains exhibit highly replicated sex differences. The most robust and extensively studied of these domains is visuospatial ability.

In 1995, Daniel Voyer, Susan Voyer, and M. P. Bryden published a definitive meta-analysis in the *Psychological Bulletin*, synthesizing decades of cognitive literature to quantify spatial sex differences[51]. They established that the magnitude of the sex difference depends intrinsically on the specific cognitive demands of the task, dividing spatial ability into three distinct sub-categories[52]:

1. **Mental Rotation:** The ability to mentally manipulate and rotate two- or three-dimensional objects in space (commonly measured by the Vandenberg-Kuse Mental Rotations Test). This task consistently shows a large male advantage, with an overall effect size of $d = 0.73$ [52]. This is one of the largest and most highly replicated cognitive sex differences in all of psychology[52].
2. **Spatial Perception:** The ability to determine spatial relationships relative to the orientation of one's own body in the presence of distracting visual information (e.g., the water-level task). This shows a moderate male advantage of $d = 0.44$ [52].
3. **Spatial Visualization:** Complex, multi-step spatial manipulations (e.g., mentally unfolding a piece of paper). This shows only a small male advantage of $d = 0.19$ [52].

To distinguish what is known about averages from what can be said about individuals, we must translate the mental rotation effect size ($d = 0.73$) into probabilities. A d of 0.73 translates to a Probability of Superiority of approximately 70%[1]. This dictates that if one man and one woman are chosen entirely at random, there is a 70% probability the man will score higher on a mental rotation task. Consequently, the overlap coefficient is approximately 69%.

Crucially, examining Cohen's U_3 reveals the proportion of women who fall outside the male average. With $d = 0.73$, approximately 23% of women will outscore the *average* (median) man[6]. Because the distributions overlap so heavily, treating spatial ability as a binary "male trait" is mathematically false. Furthermore, the extreme tails of the distribution show severe separation; roughly 4% to 5% of men will possess mental rotation scores that completely exceed the highest-scoring women in a normal distribution, and vice versa at the lower tail. It is also vital to note that spatial abilities are highly environmentally plastic and trainable; targeted spatial play and educational interventions have been shown to significantly close this gap between boys and girls[52].

Personality and Aggression

The debate surrounding sex differences in human personality is heavily polarised by methodological choices. In 2005, Janet Shibley Hyde published the "Gender Similarities Hypothesis" in the *American Psychologist*, based on a massive review of 46 meta-analyses[53]. Hyde demonstrated that for 78% of the psychological traits she examined, the sex difference was either small or near zero ($d \leq 0.35$)[53]. Based on univariate trait comparisons, Hyde

forcefully argued that men and women are overwhelmingly more alike than they are different[53]. Notable exceptions to Hyde's hypothesis of similarity include motor behaviours, specific dimensions of sexuality, and physical aggression[3]. John Archer's (2004) comprehensive meta-analysis in the *Review of General Psychology* confirmed a highly replicated, moderate-to-large male advantage in unprovoked, direct physical aggression[54]. Conversely, relational and indirect aggression (e.g., social exclusion, gossip) showed much smaller differences, sometimes favouring females[55].

However, the conclusion that male and female personalities are functionally identical was severely challenged when researchers moved from univariate trait analysis to multivariate profile analysis. In 2012, Marco Del Giudice, Tom Booth, and Paul Irwing published a landmark study in *PLoS ONE* titled "The Distance Between Mars and Venus"[2]. They argued that Hyde's methodology of examining personality traits one by one fundamentally obscures the reality that individuals possess interconnected *profiles* of traits[2].

Using a highly robust, demographically representative US validation sample of 10,261 adults who completed the 16PF Questionnaire, Del Giudice and colleagues utilised structural equation modeling to estimate sex differences on latent factors, thereby stripping away the statistical noise of measurement error[2]. When examining individual traits univariately, the differences matched Hyde's hypothesis: differences were small to moderate (e.g., Warmth $d = -0.50$, Dominance $d = 0.27$, Emotional Stability $d = 0.32$)[3].

However, when calculating the global multivariate effect size across all 15 personality axes simultaneously, the separation was exceptionally large: Mahalanobis $D = 2.71$[2]. Assuming multivariate normality, a D of 2.71 corresponds to a total overlap of only 10% between the male and female personality distributions[2]. Even when the researchers excluded the single trait with the largest univariate difference (Sensitivity) to prevent it from skewing the results, the global effect size remained massive at $D = 1.71$, which corresponds to just 24% overlap[2]. This analytical approach and its findings have since been successfully replicated by independent researchers, including a 2020 replication by Kaiser et al. using an open online dataset[56].

The implication of Mahalanobis D for the individual is profound. It dictates that while a single man and a single woman might have identical scores on a single trait like "extraversion," the mathematical probability of a man possessing a globally "feminine" profile across all 15 personality axes simultaneously is exceedingly low[2]. Personality profiles are distinctly sexually dimorphic.

The Greater Male Variability Hypothesis

Beyond differences in statistical means, the sexes differ crucially in their variance—a phenomenon formally known as the Greater Male Variability Hypothesis (GMVH). Originally noted by Charles Darwin in *The Descent of Man* and later expanded by Havelock Ellis, the

GMVH posits that males naturally display a wider probability distribution (greater variance) in many physical and cognitive traits than females[9]. Consequently, even if the male and female means are perfectly identical, greater male variance dictates that males will be vastly overrepresented at both extreme tails of the distribution—comprising the very lowest and the very highest performers[9].

Modern empirical research generally supports the GMVH in human cognition and physiology. As noted previously, Ritchie et al.'s (2018) high-powered analysis of the UK Biobank found widespread greater male variability across global and subcortical brain volumes[59]. In the realm of cognitive ability, large-scale analyses of IQ and mathematics achievement reveal Variance Ratios (VR) routinely falling between 1.08 and 1.20[9]. This indicates that male variance is 8% to 20% larger than female variance[60]. Because the tails of a distribution are hyper-sensitive to variance, a VR of 1.1 guarantees that the top 1% and bottom 1% of the cognitive distribution will be overwhelmingly male[50].

The evolutionary rationale for the GMVH is intimately tied to sexual selection. Because male reproductive success in human evolutionary history was theoretically much more skewed than female reproductive success (some high-status males sired many offspring, while many lower-status males sired none), males faced intense selective pressures to engage in high-risk, high-reward developmental and behavioural strategies[58]. Genetically, this wider variability may be mechanically facilitated by the heterogametic nature of males (possessing XY chromosomes); lacking a second X chromosome to provide genetic redundancy, recessive alleles on the X chromosome are fully expressed in males, pushing male phenotypes to both positive and negative extremes[9].

Critics of the GMVH rightly point out that the phenomenon is not a universal law of nature across all species or all traits. Harrison et al. (2022) conducted an exhaustive meta-analysis in *Biological Reviews* examining animal personality traits (boldness, aggression, exploration) across 220 species of mammals, birds, and reptiles[60]. They found absolutely no statistical evidence for greater male variability in non-human animal personality[60]. However, Del Giudice and Gangestad (2023) countered this in a published commentary, arguing that the lack of GMVH in broad, inter-species animal personality does not negate its highly robust, well-documented presence in specific human cognitive and morphological traits[60].

For human societies, the GMVH has profound implications. It suggests that exceptionally highly selective environments—such as elite mathematics institutions, extreme technical professions, or conversely, prison populations and populations experiencing severe learning disabilities—may see heavily male-skewed sex ratios. This skew is not necessarily due to bias, but is often a mathematical artifact of sampling the extreme upper and lower tails of a distribution possessing a larger variance[50].

Layer III: Culture (Variable)

The final layer of analysis addresses the environmental and societal modulation of human differences. If sex differences in personality and occupational choice are purely cultural artifacts born of patriarchal conditioning and oppressive socialisation, one would logically expect that as societies become wealthier, freer, and more egalitarian, the psychological

differences between men and women should steadily shrink. Paradoxically, the exact opposite appears to be true.

The Gender Equality Paradox

In a landmark cross-cultural study published in the *Journal of Personality and Social Psychology* in 2008, David P. Schmitt, Jüri Allik, and colleagues discovered that sex differences in Big Five personality traits were consistently *larger* in prosperous, healthy, and highly gender-egalitarian Western societies, and significantly *smaller* in poorer, less egalitarian traditional societies[61]. This highly counterintuitive phenomenon has been termed the "Gender Equality Paradox" (GEP)[62]. Subsequent high-powered research has replicated the GEP across multiple psychological domains. Falk and Hermle (2018) published a massive analysis in *Science* examining preferences for risk-taking, patience, trust, and prosociality across 76 countries with over 80,000 participants[64]. They found that gender differences in these fundamental economic preferences are strongly, positively correlated with a country's economic development and gender egalitarianism[64]. Similarly, Stoet and Geary (2018) published data in *Psychological Science* revealing that the gender gap in STEM graduation (where females are underrepresented) is actually largest in the world's most gender-equal countries (e.g., Scandinavia) and smallest in less equal, developing countries (e.g., Algeria)[63].

Competing Explanations

Two primary, highly contested mechanisms have been proposed to explain the paradox:

1. **The Evolutionary / Resource Hypothesis:** Championed by evolutionary psychologists like Schmitt, this hypothesis argues that in impoverished, oppressive, or developing societies, both men and women are forced to prioritise basic survival and economic security[61]. This intense ecological pressure forces their behaviour, personality, and career choices into similar, highly constrained patterns[61]. As societies become wealthy, safe, and egalitarian, individuals are finally freed from basic survival pressures. This unprecedented freedom acts as a catalyst, allowing men and women to fully express their innate, evolved, gender-specific preferences and biological dispositions[61].
2. **The Cultural Identity / Social Role Amplification Hypothesis:** An alternative sociological explanation argues that post-materialist, egalitarian Western societies place an incredibly high cultural premium on self-actualisation and the expression of individual identity[62]. Because gender remains a core, fundamental component of human identity, individuals in these hyper-individualistic societies may subconsciously lean into gender-typical behaviours and career paths to differentiate themselves, ironically amplifying stereotypes precisely because they are free to do so[62].

Critics of the GEP, however, suggest the entire phenomenon may be a statistical illusion. A rigorous re-analysis by Breda et al. (2024) published in the *Proceedings of the National Academy of Sciences* argued that the paradox is heavily confounded by Western cultural biases[67]. They argue it is a classic "Simpson's Paradox"; when controlling for data quality and separating countries by specific cultural regions, the correlation between gender equality and gender differentiation heavily attenuates or disappears entirely[67].

Regardless of the exact causal mechanism—whether biological unmasking or post-materialist

identity formation—the cross-cultural data definitively rule out the simplistic assumption that dismantling social barriers automatically results in men and women adopting identical psychological profiles or achieving 50/50 statistical parity in all occupations[62].

Summary and Conclusions

Robust Findings

1. The vast majority of the human population (>99.98%) is unambiguously male or female across genetic, gonadal, and genital levels; broad definitions inflating intersex prevalence to 1.7% include asymptomatic conditions (like LOCAH) and obscure physiological reality.
2. Male infant testosterone peaks massively during postnatal "minipuberty" (1-6 months, reaching 100-400 ng/dL), driving early physiological and neurological masculinisation, while female infants experience estradiol surges with negligible testosterone.
3. Within-sex variation in adult male testosterone is immense (300-1000 ng/dL), yet its direct link to individual baseline aggression is remarkably weak ($r \approx 0.08$), heavily mediated by social context, cortisol interaction, and androgen receptor (CAG repeat) genetics.
4. Human male brains are, on average, larger in volume and surface area, while female brains feature thicker cortices and higher tract complexity, though sub-regional overlap is vast.
5. The univariate "mosaic brain" theory highlights that internally consistent "extreme" brains are mathematically rare (1-6%), yet multivariate pattern recognition can accurately classify a brain's biological sex up to 93% of the time, proving robust global dimorphism.
6. General cognitive abilities show equal means between sexes, but spatial mental rotation exhibits a highly replicated, robust male advantage ($d = 0.73$, translating to a 70% probability of superiority).
7. Individual personality traits show high overlap univariately ($d \leq 0.35$), but multivariate structural equation modelling demonstrates stark overall profile separation between the sexes (Mahalanobis $D = 2.71$, resulting in roughly 10% overlap).
8. The Greater Male Variability Hypothesis is robustly supported in human cognition, mathematics achievement, and neuroanatomy, mathematically guaranteeing male overrepresentation at both the highest and lowest extremes of performance.
9. Cross-cultural research reliably demonstrates a "Gender Equality Paradox," where psychological and occupational sex differences paradoxically maximise in prosperous, highly egalitarian Western nations.
10. The constant, dynamic interaction of biological predispositions and cultural environments ensures that within-sex variation will almost always exceed average between-sex differences on any single, isolated psychological metric.

Contested or Open Questions

Domain	The Open Question	Leading Critics / Primary Debate
Endocrinology	Does cortisol reliably moderate the relationship between testosterone and status-seeking behaviour (The Dual-Hormone Hypothesis)?	Dekkers et al. (2019) and Grebe et al. (2019) point to exceptionally small effect sizes ($r = -0.061$) and severe publication bias, heavily contesting Mehta & Josephs (2010).
Neuroscience	Is the human brain inherently dimorphic or a non-binary mosaic?	Daphna Joel argues for the mosaic model based on univariate inconsistency; Marco Del Giudice counters that multivariate analysis proves undeniable dimorphism.
Sociology	What drives the Gender Equality Paradox?	David Schmitt posits that equality unmasks innate biological preferences, while Breda et al. argue it is a statistical illusion driven by Western cultural bias (Simpson's Paradox).
Evolutionary Biology	Is the Greater Male Variability Hypothesis a universal biological law applicable across species?	Harrison et al. (2022) found no evidence for it in animal personality across 220 species, a finding heavily contested by evolutionary psychologists.

Essential Sources

1. **Del Giudice, M., Booth, T., & Irwing, P. (2012).** *The distance between Mars and Venus: Measuring global sex differences in personality.* PLoS ONE. (The foundational paper establishing massive multivariate Mahalanobis D separation in human personality

profiles).
2. **Ritchie, S. J., et al. (2018).** *Sex differences in the adult human brain: Evidence from 5216 UK Biobank participants.* Cerebral Cortex. (The definitive, largest-sample macrostructural mapping of the human brain, establishing group differences and greater male variance).
3. **Joel, D., et al. (2015).** *Sex beyond the genitalia: The human brain mosaic.* Proceedings of the National Academy of Sciences. (The primary paper proposing the highly influential, yet contested, univariate mosaic brain hypothesis).
4. **Schmitt, D. P., et al. (2008).** *Why can't a man be more like a woman? Sex differences in Big Five personality traits across 55 cultures.* Journal of Personality and Social Psychology. (The foundational documentation of the Gender Equality Paradox).
5. **Voyer, D., Voyer, S., & Bryden, M. P. (1995).** *Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables.* Psychological Bulletin. (The definitive, highly replicated analysis of spatial cognition and mental rotation).
6. **Dekkers, T. J., et al. (2019).** *A meta-analytical evaluation of the dual-hormone hypothesis.* Psychoneuroendocrinology. (A rigorous, high-powered statistical critique of testosterone-behaviour links).
7. **Sax, L. (2002).** *How common is intersex? A response to Anne Fausto-Sterling.* Journal of Sex Research. (Crucial reading that clarifies the clinical prevalence of DSDs and debunks the 1.7% statistic).
8. **McGraw, K. O., & Wong, S. P. (1992).** *A common language effect size statistic.* Psychological Bulletin. (Essential reading for methodologists on how to translate Cohen's d into comprehensible probabilities of superiority).
9. **Hyde, J. S. (2005).** *The gender similarities hypothesis.* American Psychologist. (The primary, highly influential argument for univariate psychological overlap and similarity).
10. **Gettler, L. T., et al. (2011).** *Longitudinal evidence that fatherhood decreases testosterone in human males.* Proceedings of the National Academy of Sciences. (Demonstrates the profound social and developmental malleability of human male endocrinology).

Implications for Living Together

Recognising biological realities does not dictate moral imperatives. Asserting that a trait is "natural" or evolved does not mean it is "good," nor does it excuse antisocial behaviour—a cognitive trap known as the naturalistic fallacy. However, building functional, resilient partnerships, families, and societies requires engaging with human beings as they actually are, rather than as ideological blank slates.

For couples and families, the statistical data implies that enforcing rigid, absolute expectations based on biological sex is mathematically guaranteed to fail the individual. Because within-sex variation is massive across almost all univariate psychological metrics, any given man may possess lower circulating testosterone, higher emotional empathy, or lower spatial ability than his female partner. Furthermore, male hormonal architecture is deeply responsive to social bonds; the replicated drop in male testosterone upon assuming active fatherhood definitively demonstrates that male biology is profoundly adapted to nurturing and long-term paternal

investment. This directly refutes culturally constructed archetypes that demand obligatory male stoicism, detachment, or unyielding aggression.

For organisations and broader societies, the intersection of the Greater Male Variability Hypothesis and the Gender Equality Paradox provides crucial, mathematically grounded operational context. Acknowledging that males possess a wider variance in cognitive and physical traits implies that males will naturally populate the extreme failing ends of society (e.g., homelessness, incarceration, severe learning disabilities) exactly as often as they populate the extreme competitive ends. This requires societies to build tailored, compassionate social safety nets and educational interventions that account for male vulnerability at the bottom of the curve, rather than solely focusing on male overrepresentation at the top.

Furthermore, the Gender Equality Paradox suggests that organisations and governments striving for a perfect 50/50 statistical gender split in all specific occupational fields (e.g., heavy engineering, early childhood nursing) may be fighting an uphill battle against innate human preferences. The cross-cultural data indicates that in maximally free, wealthy, and egalitarian societies, men and women tend to voluntarily diverge in their career choices. Consequently, true equality of opportunity will likely result in unequal occupational outcomes. Societies can best foster human flourishing by rigorously defending equality of opportunity, ruthlessly dismantling artificial cultural barriers and biases, while remaining entirely agnostic to, and accepting of, the natural statistical disproportions that free, uncoerced individuals will subsequently create.

Works cited

1. McGraw-Wong Common Language Effect Size Calculator - MetricGate, https://metricgate.com/docs/mcgraw-wong-common-language/
2. Measuring Global Sex Differences in Personality | PLOS One, https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0029265
3. The Distance Between Mars and Venus: Measuring Global Sex, https://pmc.ncbi.nlm.nih.gov/articles/PMC3251566/
4. A Common Language Effect Size Statistic | Request PDF, https://www.researchgate.net/publication/232493691_A_Common_Language_Effect_Size_Statistic
5. A New Way to Look at the Data: Similarities Between Groups ... - PMC, https://pmc.ncbi.nlm.nih.gov/articles/PMC6428189/
6. Cohen's U3 Calculator - MetricGate, https://metricgate.com/docs/cohens-u3/
7. Heterogeneity Coefficients for Mahalanobis' D as a Multivariate, https://marcodg.net/wp-content/uploads/2018/04/delgiudice_2017_heterogeneity_d_mbr.pdf
8. 3 Relations between the standardized mean difference (Cohen's d or, https://www.researchgate.net/figure/Relations-between-the-standardized-mean-difference-Cohens-d-or-Mahalanobis-D-and_fig6_329701788
9. Variability hypothesis - Wikipedia, https://en.wikipedia.org/wiki/Variability_hypothesis
10. New Trends in Gender and Mathematics Performance: A Meta,

https://pmc.ncbi.nlm.nih.gov/articles/PMC3057475/
11. Recurring Errors in Studies of Gender Differences in Variability - MDPI, https://www.mdpi.com/2571-905X/6/2/33
12. Genetic-gonadal-genitals sex (3G-sex) and the misconception of, https://scispace.com/papers/genetic-gonadal-genitals-sex-3g-sex-and-the-misconception-of-2n2voeohky
13. How sexually dimorphic are we? Review and synthesis - ITGL, http://itgl.lu/wp-content/uploads/2015/04/Blackless-How-Dimorphic-2000.pdf
14. (PDF) How sexually dimorphic are we? Review and synthesis, https://www.researchgate.net/publication/11812321_How_sexually_dimorphic_are_we_Review_and_synthesis
15. How Common is Intersex? An Explanation of the Stats., https://www.intersexequality.com/how-common-is-intersex-in-humans/
16. Intersexuality - wikidoc, https://www.wikidoc.org/index.php/Intersexuality
17. How common is Intersex? A response to Anne Fausto-Sterling, https://www.researchgate.net/publication/10996431_How_common_is_Intersex_A_response_to_Anne_Fausto-Sterling
18. 1.7% of This Episode Is Intersex - Apple Podcasts, https://podcasts.apple.com/ie/podcast/1-7-of-this-episode-is-intersex/id1784533693?i=1000774457931
19. Up-To-Date Review About Minipuberty and Overview on ... - PMC, https://pmc.ncbi.nlm.nih.gov/articles/PMC6070773/
20. Mini puberty and its interpretation - PMC, https://pmc.ncbi.nlm.nih.gov/articles/PMC4462291/
21. (PDF) Meta-analysis of digit ratio 2D:4D shows greater sex, https://www.researchgate.net/publication/45828808_Meta-analysis_of_digit_ratio_2D4D_shows_greater_sex_difference_in_the_right_hand
22. Congenital hypogonadotropic hypogonadism, https://www.pfmjournal.org/journal/view.php?doi=10.23838/pfm.2021.00093
23. Decoding Mini-Puberty and Its Clinical Significance: A Narrative, https://www.mdpi.com/2673-396X/6/2/28
24. Minipuberty: Looking Back to Understand Moving Forward - PMC, https://pmc.ncbi.nlm.nih.gov/articles/PMC7848193/
25. The Course of Minipuberty in Daughters of Women with Low ... - MDPI, https://www.mdpi.com/2072-6643/16/14/2362
26. A meta-analytical evaluation of the dual-hormone hypothesis, https://www.researchgate.net/publication/329845235_A_meta-analytical_evaluation_of_the_dual-hormone_hypothesis_Does_cortisol_moderate_the_relationship_between_testosterone_and_status_dominance_risk_taking_aggression_and_psychopathy
27. The Relationship Between Testosterone and Aggression: A Meta, https://www.researchgate.net/publication/222323023_The_Relationship_Between_Testosterone_and_Aggression_A_Meta-Analysis
28. Testosterone and human behavior: the role of individual and, https://wpms.nipissingu.ca/snelab/wp-content/uploads/sites/32/2014/10/Carre-Ar

cher-2017.pdf
29. The dual-hormone hypothesis and first-time fathers' relationship, https://pmc.ncbi.nlm.nih.gov/articles/PMC12137339/
30. Testosterone, cortis - IRIS-AperTO, https://iris.unito.it/retrieve/handle/2318/1853339/974937/Grebe_etal_2019_dual-hormone-hypothesis_pre.pdf
31. Beyond the Challenge Hypothesis: The Emergence of the Dual, https://pmc.ncbi.nlm.nih.gov/articles/PMC7311256/
32. A preliminary case study of androgen receptor gene polymorphism, https://www.dovepress.com/a-preliminary-case-study-of-androgen-receptor-gene-polymorphism-associ-peer-reviewed-fulltext-article-AGG
33. Androgen receptor (AR)-CAG trinucleotide repeat length and ... - PMC, https://pmc.ncbi.nlm.nih.gov/articles/PMC6341423/
34. Size Matters: The CAG Repeat Length of the Androgen Receptor, https://www.frontiersin.org/journals/psychiatry/articles/10.3389/fpsyt.2021.732759/full
35. Androgen Receptor Gene CAG Repeat Length and Body Mass, https://academic.oup.com/jcem/article/92/10/3844/2598090
36. Testosterone, androgen receptor gene CAG repeat length ... - PubMed, https://pubmed.ncbi.nlm.nih.gov/20479013/
37. A meta-analysis of the association between male dimorphism ... - eLife, https://elifesciences.org/articles/65031
38. Testosterone Levels Are Negatively Associated with Childlessness, https://pmc.ncbi.nlm.nih.gov/articles/PMC3616053/
39. No evidence for relationship between paternal post-partum, https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2024.1348031/full
40. A pilot prospective cohort investigating paternal hormonal changes, https://www.tandfonline.com/doi/full/10.1080/02646838.2025.2523340
41. Sex Differences in the Adult Human Brain: Evidence from 5216 UK, https://academic.oup.com/cercor/article-abstract/28/8/2959/4996558
42. Sex differences in the adult human brain - Sign in, https://www.pure.ed.ac.uk/ws/files/59812674/bhy109.pdf
43. Sex Differences in the Adult Human Brain: Evidence from 5216 UK, https://scispace.com/pdf/sex-differences-in-the-adult-human-brain-evidence-from-5216-1umpqu9ja9.pdf
44. Are Sex Differences in Human Brain Structure Associated ... - PMC, https://pmc.ncbi.nlm.nih.gov/articles/PMC8726594/
45. Greater male than female variability in regional brain structure, https://research.rug.nl/files/146910058/hbm.25204.pdf
46. Sex beyond the genitalia: The human brain mosaic - PNAS, https://www.pnas.org/doi/10.1073/pnas.1509654112
47. Sex beyond the genitalia: The human brain mosaic - PNAS, https://www.pnas.org/doi/abs/10.1073/pnas.1509654112
48. Joel et al.'s method systematically fails to detect large, consistent sex,

https://marcodg.net/wp-content/uploads/2016/03/delgiudice_etal_2016_letter_joel-etal_pnas.pdf
49. Analysis of Human Brain Structure Reveals that the Brain “Types, https://pmc.ncbi.nlm.nih.gov/articles/PMC6204758/
50. The Greater Male Variability Hypothesis - Heterodox Academy, https://heterodoxacademy.org/blog/the-greater-male-variability-hypothesis/
51. Mental rotation performance and gender-roles - a meta-analysis, https://research-repository.griffith.edu.au/bitstreams/6846a263-fcfc-591c-b09e-9cc923060d90/download
52. Spatial Reasoning: Gender Differences - JobCannon, https://jobcannon.io/blog/spatial-reasoning-gender-differences
53. Gender Similarities and Differences : Annual Review of Psychology, https://www.ovid.com/journals/anrp/fulltext/10.1146/annurev-psych-010213-115057~gender-similarities-and-differences
54. Sex differences in adolescent physical aggression, https://dspace.library.uu.nl/server/api/core/bitstreams/767032b3-2735-4653-825e-ecbe47089f1e/content
55. Archer, J., & Mehdikhani, M. 2004. Sex Differences in Aggression in, https://www.scribd.com/document/416237856/Archer-J-Mehdikhani-M-2004-Sex-Differences-in-Aggression-in-Realworld-Settings-a-Metanalytic-Review-Review-of-General-Psychology-8-291-236
56. Global sex differences in personality: Replication with an open, https://www.researchgate.net/publication/334279951_Global_sex_differences_in_personality_Replication_with_an_open_online_dataset
57. Publikationsserver / Global sex differences in personality, https://eplus.uni-salzburg.at/obvusboa/content/titleinfo/4605464
58. An Evolutionary Theory for the Variability Hypothesis - arXiv, https://arxiv.org/html/1703.04184v15
59. Sex Differences in Variability of Brain Structure Across the Lifespan, https://pmc.ncbi.nlm.nih.gov/articles/PMC7566684/
60. No Evidence Against the Greater Male Variability Hypothesis - ArTS, https://arts.units.it/retrieve/b83d3703-73ef-4e86-af8a-46cee702b6ac/DelGiudice_Gangestad_2023_Commentary_Harrison_pre_v4.pdf
61. "Why Can't a Man Be More Like a Woman? Sex Differences in Big, https://www.researchgate.net/publication/24001221_Why_Can't_a_Man_Be_More_Like_a_Woman_Sex_Differences_in_Big_Five_Personality_Traits_Across_55_Cultures_vol_94_pg_168_2008
62. Gender-equality paradox in competitiveness: Evidence and, https://davs-econ.github.io/competitiveness_pisa_full.pdf
63. Linking gender differences with gender equality: A systematic ... - PMC, https://pmc.ncbi.nlm.nih.gov/articles/PMC9978710/
64. Gender Differences in the Intention to Start a Business, https://econtent.hogrefe.com/doi/10.1027/2151-2604/a000435
65. Gender differences in cooperation across 20 societies: a meta, https://royalsocietypublishing.org/rstb/article/378/1868/20210438/42662/Gender-

differences-in-cooperation-across-20
66. The Gender Gap in the Care Economy Is Larger in... - LWW.com, https://journals.lww.com/00000487-202603000-00008
67. Simpson's gender-equality paradox - PNAS, https://www.pnas.org/doi/10.1073/pnas.2422247122