MASTERARBEIT | MASTER'S THESIS

Titel | Title

Beauty and the Genes: A multiverse meta-analysis of the Association between Intelligence and Attractiveness

verfasst von | submitted by

Petra Tauchner BSc BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 840
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Psychologie
Betreut von | Supervisor:	Assoz. Prof. Mag. Dr. Jakob Pietschnig Privatdoz.

Danksagung

Zunächst möchte ich meinem Betreuer Ass.-Prof. Mag. Dr. Jakob Pietschnig danken. Ihre Leidenschaft und Begeisterung für die Wissenschaft inspirierten mich seit der ersten TEWA-Vorlesung und motivierten mich, ebenso viel Zeit und Energie in meine Masterarbeit zu investieren. Ihr kritisches Denken und Ihre unermüdliche Neugier bereichern unsere Gesellschaft. Danke, dass Sie trotz unzähliger Projekte dennoch Zeit für persönliche Gespräche mit all Ihren Studenten finden.

Weiteren Dank möchte ich Sandra aussprechen. Ohne deinen Zuspruch hätte ich mich nie an dieses riesige Thema gewagt. Du bist eine der intelligentesten und stärksten Frauen, die ich kennenlernen durfte. Deine von Grund auf herzliche und hilfsbereite Persönlichkeit hat mich schon nach unserem ersten Gespräch in den Bann gezogen. Danke, dass du nie aufgibst und dich stets neuen Herausforderungen stellst und dich dabei immer wieder selbst übertriffst – eine *richtige* Power-Frau eben.

Von ganzem Herzen möchte ich Kerstin danken. Ohne deine Unterstützung, sei es bei Präsentationen oder beim alltäglichen Austausch, wäre ich nicht dort, wo ich jetzt bin. Unsere langjährige, einzigartige Freundschaft gibt mir immens viel Kraft, uns verbindet mehr als Worte es je ausdrücken könnten. Deine Loyalität, dein Zuspruch und unser gemeinsames Wachstum bereichern mein Leben in Ausmaßen, die ich kaum fassen kann. Mit dir an meiner Seite lässt sich jedes noch so große Hindernis bewältigen. Du bist immer für mich da, egal mit welchen Aufgaben uns das Leben konfrontiert, und ebenso möchte ich für dich da sein.

Weiters möchte ich Marlene und Michi danken. Ihr habt mich zu Höchstleistungen gepusht, euer Ehrgeiz und persönlicher Erfolg sind eine riesige Inspiration. Wir sind gemeinsam durch schwierige Zeiten gegangen und ich bin endlos dankbar, mein Leben mit euch zu teilen – und vize-versa.

Vor allem danke ich auch Arthur. Du hast seit Beginn unseres Kennenlernens an mich geglaubt und etwas in mir gesehen, was für mich zu dieser Zeit noch verschleiert war. Durch dich habe ich mich sowohl persönlich als auch zwischenmenschlich enorm weiterentwickelt. Danke für all die Diskussionen, Erlebnisse und Lernerfahrungen. Ich schätze deine Einstellung zum Leben, deinen Drive und vor allem deine einzigartige Seele.

Zuletzt danke ich meinem Bruder. Unsere Zeit während des Biologiestudiums, vor allem die gemeinsame Arbeit im Labor, hat mich erst so richtig erkennen lassen, was für ein intelligenter, wissbegieriger und hartnäckiger Mensch du bist. Dein moralischer Kompass hat

mich in meiner Jugendzeit mutmaßlich geprägt. Du warst ein Vaterersatz, ein Vorbild und vor allem ein guter Freund. Uns verbindet weit mehr als nur Blut, Schweiß und Tränen. Ich weiß, wir sprechen es nie persönlich aus, aber *ich liebe dich von ganzem Herzen* und bin unendlich stolz und froh, dich meinen Bruder nennen zu dürfen. Danke für deine Unterstützung, danke für all die Gespräche und danke fürs gemeinsame Lachen, in schlechten als auch guten Zeiten.

Ein Kapitel meines Lebens schließt sich nun, und ich blicke mit Zuversicht in die Zukunft mit all ihren grenzenlosen Möglichkeiten und bereichernden Erfahrungen. Um abschließend meinen Lieblingsautor zu zitieren: *„Jedem Anfang wohnt ein Zauber inne, der uns beschützt und der uns hilft, zu leben.“* – Hermann Hesse.

Ihr bereichert mein Leben, schenkt mir Kraft, lässt mich wachsen und verleiht mir unaufhaltsamen Optimismus.

Danke.

Content

Abstract 1
Introduction 2
Intelligence 2
Facial Symmetry 3
Averageness 4
Sexual Dimorphism 5
The Physical Attractiveness Stereotype 6
On the Attractiveness-Intelligence Association: Updating Evidence from Kanazawa & Kovar (2004) 9
Evidence from prior Meta-Analyses on the Attractiveness-Intelligence Association 16
Methods 16
Literature search 17
Inclusion criteria 17
Coding 17
Data analysis 18
Final sample 23
Results 26
Main analyses 26
Subgroup analyses 26
Moderator analysis 27
Dissemination Bias 32
Specification curve analysis 35
Combinatorial meta-analysis 36
Discussion 39
The Dual Influence of Attractiveness: About the Health-Attractiveness Association, Social Reinforcement, and its Impact on Intelligence 39
The Declining Correlation Between Attractiveness and Intelligence Over Time 41

Age-Dependent Differences in the Attractiveness-Intelligence Correlation: Stronger Association in Children Compared to Adults 41

Examining the Influence of Live and Photograph-Based Attractiveness Assessments on the Relationship Between Intelligence and Attractiveness 43

Publication Bias Analysis Corroborates the Validity of the Attractiveness-Intelligence Relationship 44

The Stability of the Attractiveness-Intelligence Relationship: Findings from Specification Curve Analysis and Combinatorial Meta-Analyses 44

Limitations 45

Conclusion 46

References 50

List of Figures 71

List of Tables 72

Appendix A 73

Appendix B 75

Appendix C 76

Appendix D 79

Appendix E 81

References Appendix E 86

Appendix F 89

References Appendix F 91

Appendix G 93

References Appendix G 97

Appendix H 98

Appendix I 119

Appendix J 120

Abbreviations 123

Abstract

The correlation between physical attractiveness and intelligence has long intrigued evolutionary theorists. One prominent explanation is based on cross-trait assortative mating (xAM), where individuals with high levels of one desirable trait tend to pair with those who have high levels of another desirable trait. Two large meta-analyses have shown a small positive correlation between physical attractiveness and intelligence. However, prior meta-analyses have incorporated studies utilizing measures such as grade point averages, school performance, and academic achievement to evaluate intelligence, potentially biasing the outcomes. The present meta-analysis is the first to exclusively analyze studies that used psychometric IQ-tests $n = 9$ ($k = 14$, $N = 50{,}156$). Random-effects analyses showed a small yet robust positive association between physical attractiveness and intelligence $r = 0.13$ ($p = .0005$, 95% CI [0.06, 0.20]). Even after removing an outlier study that showed an exceptionally high correlation, the association between physical attractiveness and intelligence remained robust albeit slightly smaller $r = 0.10$ ($p = .0004$, 95% CI [0.04, 0.15]). No strong evidence was found for publication bias. Moreover, combinatorial analyses provided further support for the robustness of the results. Overall, the results indicate a small and robust association between physical attractiveness and intelligence.

Keywords: Intelligence, physical attractiveness, cross-trait assortative mating, meta-analysis, multiverse-analysis.

Introduction

From the dawn of human existence, beauty has captivated our minds. Despite longstanding beliefs in the subjectivity of attractiveness, scientific research over the past decades has shown that there are certain facial and bodily characteristics that are recognized as universally beautiful. Multiple studies provide evidence of high agreement in attractiveness ratings both within and across diverse cultures and ethnicities (e.g., Cunningham et al., 1995; Langlois et al., 2000; Marcus & Miller, 2003; Coetzee et al., 2014; Holzleitner et al., 2019). While individual preferences can vary, the underlying principles of symmetry, proportion, and certain health indicators remain constant. This universality of beauty not only highlights the commonalities in human perception but also underscores the potential evolutionary advantages of these preferences.

Intelligence

In this thesis, the discussion will be confined to psychometric or academic intelligence, respectively. Although many models of intelligence were established throughout the past decades, one of the prevailing models today is the *Cattell-Horn-Carroll* (CHC) model of intelligence (see Schneider & McGrew, 2012). According to the CHC, intelligence constitutes a hierarchical structure of cognitive abilities encompassing narrow, broad, and general abilities with general intelligence (factor *g*) influencing cognitive skills. The existence of psychometric *g* is well-supported and widely accepted in the scientific community. For example, a meta-analysis conducted by Zaboski II et al. (2018) revealed that the general cognitive factor (*g*) accounted for a greater proportion of the variation in academic achievement compared to the cumulative influence of all broad cognitive abilities considered together (mean effect size of $r^2 = 0.540$). Estimations on the heritability of intelligence commonly range between 20% to 80% (e.g., Plomin & Spinath, 2004; Petrill et al., 2004; Haworth et al, 2010; Plomin & Deary, 2015), with estimations increasing over the lifespan of individuals. Genome-wide association analyses have also found numerous genomic loci and genes linking to intelligence (Savage et al., 2018). While it remains challenging to precisely quantify the extent of genetic influence, the evidence strongly suggests a substantial role of genetics.

Facial Symmetry

Symmetric faces and objects are generally regarded as more beautiful (Grammer & Thornhill, 1994; Gangestad et al., 1994; Perrett et al., 1999; Scheib et al., 1999; Bertamini et al., 2019). This preference for symmetry is observed even in children as young as four years old, who display a natural inclination towards symmetric objects (Huang et al., 2018). However, it is noteworthy that while symmetry contributes to attractiveness, it may only offer a partial explanation for facial appeal (Zaidel et al., 2005; Jones et al., 2007). For instance, recent research suggests that although facial symmetry is an integral component, it is not a direct determinant of facial attractiveness. Specifically, there appears to be an interactive effect between symmetry and perceived normality, wherein perceived normality acts as a mediator between symmetry and facial attractiveness. When facial symmetry decreases perceived normality, faces are rated as less attractive. Conversely, when faces are perceived as more familiar, symmetry increases attractiveness (Zheng et al., 2021). Zaidel & Hessamian (2010) examined facial attractiveness of half faces. Three groups of participants rated attractiveness of full-faces and half faces (vertical left and right hemi-faces). No significant differences were found for attractiveness ratings of full- vs. hemi-faces. Rather, they found a strong positive correlation between ratings of semi- and full-faces, indicating that bilateral symmetry is not a pivotal factor in attractiveness assessment. Humans seem to prefer natural faces with some degree of asymmetry, at least in comparison to perfect mirror images (Choi, 2015).

One common explanation for the attractiveness of symmetry revolves around the subconscious recognition of fluctuating asymmetry in facial features. Fluctuating asymmetry (i.e., deviations from bilateral symmetry) is linked to pathogen exposure or other toxins during development, indicating developmental instability. Such asymmetry tends to increase with factors like inbreeding, bad nutrition, and environmental pollutants (Møller & Thornhill, 1998; Graham & Özener, 2016). Symmetric faces therefore represent high genetic quality and low pathogen exposure during development. A meta-analysis investigating the effects of fluctuating asymmetry on mating success revealed that there is a large overall negative association between mating success and fluctuating asymmetry ($r = -0.42$, $p < .0005$), with this effect being more pronounced in men than women (Møller & Thornhill, 1998). This negative relationship is often interpreted as evidence that sexual selection favors offspring with higher genetic quality (i.e., greater potential to fight off parasites, representing higher heterozygosity). Fluctuating asymmetry has also been linked to lower female fecundity and increased parasite exposure (Thornhill & Gangestad, 2006; Møller, 1998). Additionally, there is evidence suggesting that symmetric faces may be perceived as more beautiful because they are processed more easily by the visual system. Infants as young as 12 months old show a preference for facial symmetry (Griffey & Little, 2014). This observation may suggest the presence of an intrinsic mechanism for facial attractiveness perception. Research by Trujillo et al. (2014) further supports the hypothesis of processing fluency for attractiveness. In their study, participants had to categorize attractive human faces, unattractive human faces and chimpanzee faces while event-related potentials (ERPs) and reaction times were recorded. For average and high-attractive faces, categorization to “human” was faster, indicating that attractive faces are processed more easily. Recognizing these underlying fitness indicators has been crucial for human reproduction and are integral to mate selection processes. This innate preference for symmetry aligns with evolutionary principles of selecting partners with traits indicative of robust health and genetic quality, ultimately enhancing the chances of reproductive success.

Averageness

Numerous experiments have shown that people rate average faces as more attractive compared to non-average faces (Langlois & Roggman, 1990; Rhodes & Tremewan, 1996; Valentine et al., 2004; Rhodes et al., 2007). The parasite theory of sexual selection offers a compelling explanation for this phenomenon, suggesting that individuals find average faces attractive because they reflect high individual protein heterozygosity (i.e., genetic diversity), thereby enhancing pathogen resistance (Thornhill & Gangestad, 1993). Additionally, average faces are associated with functional advantages, such as optimal features for breathing and chewing (Thornhill & Gangestad, 1999; Rhodes, 2006). Consequently, averageness is linked to reproductive success and mate quality. Within an evolutionary context, the adaptationist framework plays a pivotal role in understanding the development of attractiveness and beauty. The adaptationist framework seeks to identify selective pressures that have shaped an organism’s evolutionary development (Gangestad & Scheyd, 2005, p. 525). The emergence of attractiveness and beauty is conceptualized as an outcome of adaptation, with individuals naturally drawn to others possessing specific favored traits (e.g., averageness). In the context of the adaptationist framework, the prevailing hypothesis posits that these perceiver traits are adaptations themselves, evolved due to their inherent benefits for perceivers (e.g., average faces to display high heterozygosity or functionality).

Sexual Dimorphism

Facial masculinity (e.g., strong jaw and cheek bones, prominent brow ridge, facial hair) has been linked to higher testosterone levels and immunological competence (see Pound et al., 2009; Rantala et al., 2012; Boothroyd et al., 2013; Trigunaite et al., 2015). Interestingly, some research suggests that higher testosterone levels impair immunocompetence (Fuxjager et al., 2011; Gubbels Bupp & Jorgensen, 2018). According to the immunocompetence handicap hypothesis (ICHH), it is assumed that the development and maintenance of strong masculine traits are energetically costly and, only individuals with a strong immune system can afford to invest resources in such characteristics without compromising their overall health. Thus, prominent facial masculine features serve as a fitness indicator for females. However, evidence on the ICCH remains equivocal. A recent study investigated the effects of testosterone levels on the effectiveness of the influenza vaccine and found no evidence of immunosuppression in male humans with higher testosterone and androgen levels. Moreover, the results revealed that the strength of the immune response to the influenza vaccine was positively associated with androgen markers (Nowak et al., 2018). Studies investigating the ICHH often lean towards observation rather than experimentation, leading to challenges in result interpretation. A meta-analysis by Roberts et al. (2004) also found a significant effect of testosterone on immune suppression. However, the effect disappeared when controlling for multiple studies within the same species. As of now, evidence surrounding the validity of the ICHH remains unclear and must be further tested in humans.

In women, high estrogen levels are associated with reproductive fitness (Smith et al., 2006; Smith et al., 2012). Beyond its direct impact on reproductive capabilities, estrogen may also influence maternal tendencies (i.e., the desire to have children). Indeed, a positive relationship between estrogen levels and the reported number of desired children was found. Women showing higher estrogen levels also tend to have more feminized faces (e.g., fuller lips, smaller chin and nose, big eyes) compared to those who desire less children (Smith et al., 2012). Additionally, estrogen plays a crucial role in maintaining skin health since it has protective effects on soft tissue. Loss of estrogen in women over time has been linked to skin aging (Lephart, 2018). Therefore, lower estrogen levels can accelerate the appearance of aging signs such as wrinkles and decreased skin elasticity, potentially influencing perceptions of youthfulness and overall attractiveness. Fluctuating estrogen levels during the menstrual cycle also influence mate choice. Concealed ovulation is a distinctive characteristic of humans, setting us apart from many other species where ovulation is more apparent. In many non-human primates, for example, females exhibit clear physical signs of fertility, such as genital swelling, which signal their reproductive status to males. However, in humans, there are no overt indicators of ovulation, making it difficult for others to detect when a female is most fertile (Buss, 2019). Interestingly, subtle changes do occur during the ovulatory phase, such as slight enhancements in lip and skin coloration, which can increase the attractiveness of female faces. Both men and women rate photographs as more attractive when taken during the ovulatory phase, indicating high estrogen levels (Roberts et al., 2004; Puts et al., 2013). This subtlety may have evolved to promote pair bonding and paternal investment by obscuring the exact timing of fertility, thus encouraging males to remain with their partners throughout the reproductive cycle. A meta-analysis exploring women's mate preferences during the ovulatory cycle discovered that naturally fluctuating hormone levels also affect women's mate choices for short-term vs. long-term attractiveness of men (Gildersleeve et al., 2014). Women exhibited stronger preferences for characteristics displaying genetic quality in males during the ovulation phase compared to low-fertility days (luteal phase). Importantly, this shifting effect in mate preference was dependent on the type of relationship (long-term vs. short-term). The effect was more prominent when evaluating the mate quality for a short-term relationship (e.g., a one-night stand), whereas no effects were found in terms of choices for long-term relationships (e.g., marriage).

Both more masculine faces in men and more feminine faces in women are typically rated as more attractive (Gangestad & Thornhill, 2003; Smith et al., 2012; Lephart, 2018). When modifying the degree of sexual dimorphism in the face digitally, humans cross-culturally prefer feminized faces in women and masculinized faces in men. Furthermore, masculinized faces are rated as more dominant and less dishonest, indicating the multifaceted nature of facial preferences across different cultures and contexts (Perrett et al., 1998).

The Physical Attractiveness Stereotype

The physical attractiveness stereotype describes the tendency of people to associate positive traits such as intelligence, competence, friendliness, and social skills with individuals who are perceived as physically attractive (Dion et al., 1972). Recent research continues to affirm the prevalence of the physical attractiveness stereotype. For instance, Lorenzo et al. (2010) employed a round-robin design wherein participants individually met with each other for brief three-minute intervals. Following assessments of attractiveness, participants also rated each other using a 21-item version of the Big-Five Inventory (BFI) and on three items gauging intelligence. The results revealed that within this short interaction period, physically attractive individuals were rated more positively and more accurately. In line with the physical attractiveness stereotype, more attractive individuals were perceived as possessing more positive characteristics. Moreover, attractiveness influences everyday contexts, such as tourism and customer service. As an example, in hospitality settings, attractive guests tend to receive more positive evaluations from hotel employees regarding their propensity to spend money, their levels of demand, and their kindness (Čivre et al., 2013). Similarly, research revealed that customer satisfaction, perceptions of service quality, and likability of service representatives were all enhanced when those representatives were deemed more attractive (Li et al., 2019). Further confirmation of societal bias was found in a recent study, which showed that participants rated attractive candidates more favorably in terms of hiring and interpersonal compatibility (Chance et al., 2023).

Nonetheless, some research also indicates a downside to the attractiveness stereotype, particularly concerning its impact on women and members of the opposite sex. Heilman & Stopeck (1985) first coined the term "beauty is beastly" to elucidate the adverse consequences experienced by attractive women seeking jobs in traditionally masculine roles. Building upon this notion, subsequent empirical investigations have explored the "beauty is beastly hypothesis". The results show a nuanced pattern: while attractiveness conferred advantages for men irrespective of job type, for women, attractiveness proved to be more advantageous in positions traditionally associated with feminine stereotypes compared to those aligned with masculine stereotypes (Johnson et al., 2010). Analysis of a large dataset of over 20,000 individuals also showed that there was no positive association between physical attractiveness and earnings. Instead, it is suggested that individual differences in health, intelligence and personality play a more substantial role in determining earnings (Kanazawa & Still, 2018).

As described by the modern term "pretty privilege", the attractiveness stereotype not only refers to positive character traits, but also preferential treatment by society in large. The phenomenon of the attractiveness stereotype appears to be a globally pervasive concept. Generally, research indicates that individuals tend to exhibit favorable biases towards attractive individuals of the opposite sex, whereas their attitudes towards attractive individuals of the same sex are less positive, sometimes even displaying negative biases as discussed before (Agthe et al., 2010; Watkins et al., 2010; Arnocky et al., 2012). It was further shown that participants from diverse ethnic backgrounds displayed similar trends showing positive biases towards attractive individuals of the opposite sex and negative biases toward attractive individuals of the same sex within their respective ethnic groups (Agthe et al., 2016). This suggests a universality of the attractiveness bias across different cultures, shedding light on intrasexual competition strategies. Results from a comprehensive meta-analysis also supported the prevalence of a robust attractiveness stereotype. Both children and adults receive more positive treatments from strangers as well as by people in their close environment. Moreover, attractive individuals display more positive behaviors (i.e., higher extraversion, self-esteem, social skills, career success and dating experience) (Langlois et al., 2000). A recent analysis updating evidence for the attractiveness stereotype was conducted using data from 45 countries and over 11,000 subjects. Participants were asked to rate 120 faces on 13 adjectives, including attractiveness, aggression, caringness, confidence, dominance, emotional stability, intelligence, meanness, responsibility, sociability, trustworthiness, unhappiness, and weirdness. Results showed that attractiveness correlated positively with all socially desirable traits and negatively with socially undesirable traits. Both for female and male faces and across different countries, attractive faces correlated positively with socially desirable personality traits (Bartres & Shiramizu, 2023). Using computational models for varying individual's faces (e.g., face shape, skin smoothness, and pigmentation), it was shown that these models accurately predict human social judgments such as attractiveness, competence, dominance, and trustworthiness, and can differentiate between these judgments despite their similarities (Todorov et al., 2013). These attributes make attractiveness a well-defined and consistent stereotype. The models' ability to differentiate between various social traits suggests that attractiveness is part of a broader set of social stereotypes influencing our perceptions and interactions. Moreover, the stability and predictability of these judgments indicate that attractiveness-related biases are deeply ingrained and pervasive, affecting social interactions and potentially leading to advantages for those deemed attractive. Longitudinal studies have also shown that the positive biases associated with attractiveness remain consistent as individuals age, indicating that these perceptions are not merely transient but are deeply ingrained in social cognition (Zebrowitz et al., 1998; Zebrowitz et al., 2013). Additionally, cross-cultural research consistently finds that despite slight variations in beauty standards, the core attributes linked to attractiveness, such as health, confidence, and sociability are universally valued (Jones & Hill, 1993; Langlois et al., 2000). This enduring nature of the attractiveness stereotype underscores its significant role in shaping social dynamics and individual interactions across diverse societies.

On the Attractiveness-Intelligence Association: Updating Evidence from Kanazawa & Kovar (2004)

Humans tend to rate more attractive individuals as more intelligent, however, it is unclear whether perceived intelligence reflects actual intelligence. Four assumptions were proposed to explain the correlation between IQ and attractiveness.

1. *More intelligent men are more likely to occupy higher status than less intelligent men.*
2. *Higher-status men are more likely to mate with more beautiful women than lower-status men.*
3. *Intelligence is heritable, such that sons and daughters of more intelligent men are more intelligent than sons and daughters of less intelligent men.*
4. *Beauty is heritable, such that sons and daughters of more beautiful women are more beautiful than sons and daughters of less beautiful women.*

If all assumptions are empirically true and observable, the conclusion that attractive people are more intelligent must be logically true. The link between IQ and physical attractiveness can be explained through the concept of cross-trait assortative mating (xAM). xAM describes an individual's tendency to mate with partners who have a specific, different phenotype (e.g., physical attractiveness, higher levels of dominance etc.), which usually have no genetic relationship (Border et al., 2022). Individuals select mates based on particular characteristics or fitness markers. xAM is based on the concept that traits that are, per se, unrelated (e.g., height and wealth) become jointly inherited. In the context of attractiveness and intelligence, it is posited that these traits are inherited together due to underlying mechanisms or indications of fitness (i.e., intelligent males choosing high-status jobs and mating with more attractive women), thus steering sexual selection in this direction.

"More intelligent men are more likely to occupy higher status than less intelligent men."

A recent study reinforced a positive correlation between intelligence and occupational status, suggesting individuals with higher intelligence tend to achieve higher positions or status in society. A positive relationship was found between IQ and job performance as well as job satisfaction, which was mediated by personality type (A/B) (Murtza et al., 2021). Moreover, the results from a comprehensive meta-analysis further underscore the intricate interplay between IQ, socioeconomic status (SES) and various indicators of success such as education, occupation, and income. While the results demonstrate that IQ is a strong predictor of success, it is overall not a superior predictor compared to parental SES or grades (Strenze, 2007). However, these causality questions are complex due to multiple covariations. Three causal models can be used to explain the intricate relationship between SES, intelligence and academic achievement, i) one explanation positing independent influence of intelligence and SES on academic achievement, ii) another suggesting intelligence solely as a proxy for SES, and iii) a third model proposing that SES has no direct influence on academic achievement beyond its impact on intelligence (Hunt, 2010). Remarkably, the third model emerges as the most accurate representation, indicating that while intelligence is indeed a predictor of academic achievement, its relationship with SES plays a pivotal role in understanding academic success.

In support of the significant role of intellectual ability, a longitudinal study conducted on a Swedish sample of over a thousand participants found that the likelihood of achieving at least a master's degree was significantly higher (almost 10 times more common) for individuals belonging to the top 10% of population in terms of intellectual ability (IQ score of 120 or above) compared to participants with an average IQ (Bergman et al., 2014). Additionally, research has shown that cognitive ability not only predicts job performance but also career success over time, with individuals of higher IQ being more likely to experience upward mobility in their careers (Judge et al., 2004). Furthermore, intelligence measured in childhood is a significant predictor of occupational status in adulthood, independent of parental SES (Čukić et al., 2017).

While numerous studies suggest a positive correlation between intelligence and occupational status, it is essential to consider evidence highlighting the limitations and complexities of this relationship. For instance, non-cognitive factors such as social skills, motivation, and personality traits significantly influence occupational success, sometimes even surpassing the impact of cognitive abilities (Heckman et al., 2006). Moreover, other factors such as emotional intelligence and social skills have a large impact on career advancement. Individuals with high emotional intelligence are better equipped to navigate social complexities in the workplace, thereby achieving higher status (O'Boyle et al., 2011). Furthermore, research contends that economic and social capital, which includes networking and family background, plays a crucial role in determining occupational outcomes, often overshadowing the influence of IQ (Deming, 2017). These findings indicate that while intelligence is a factor in achieving higher occupational status, it is far from the sole determinant, and other variables such as social skills, emotional intelligence, and social capital are equally, if not more, important in predicting career success.

"Higher-status men are more likely to mate with more beautiful women than lower-status men."

From an evolutionary perspective, resources and status are indicators of a male's ability to provide for offspring. This is based on the principle of parental investment (Trivers, 1972), which suggests that the sex making the largest investment in offspring (which is typically the female in mammals due to gestation and nursing) will be more selective in choosing a mate. A substantial body of research suggests gender differences in mate selection, interest in casual sex, jealousy, and the desire for sex variety (Buss, 1989; Petersen & Hyde, Baumeister et al., 2001; Guadagno & Sagarin, 2010; Petersen & Hyde, 2010). These differences range from small to large effects, with differences in casual sex and mate selection being more pronounced. While men tend to put more importance on physical attractiveness and fecundity, women tend to value characteristics such as socioeconomic status (SES) and industriousness, reflecting a preference for traits associated with resource acquisition and provision (Buss, 1989; Feingold, 1992; Gangestad & Simpson, 2000; Regan et al., 2000; Li et al., 2011; for a review see Buss & Schmitt, 2019). These gender differences are also more pronounced in short-term relationships and become more modest in long-term relationships (Schmitt, 2005). For long-term relationships, both men and women prioritize internal traits like kindness, reliability, warmth, and intelligence (Buss, 1989; Regan et al., 2000; Eastwick & Finkel, 2008).

According to Buss and Schmitt's Sexual Strategies Theory (1993), men and women have evolved different mating strategies. Men who can produce numerous offspring with minimal investment, are more likely to seek out physically attractive mates as a sign of fertility and hence pursue short-term mating strategies more often than women. The males-compete/females-choose (MCFC) model proclaims that males primarily compete with each other for access to females, while females are more selective in choosing their mates. This model is in line with the Sexual Strategies Theory, suggesting that men have evolved to pursue short-term sexual relationships with multiple partners, aiming for quantity over quality in mating, while women evolved to be choosier about their sexual partners and to prefer long-term pair bonding with men who can provide for their offspring. In contrast, the mutual mate choice (MMC) model suggests that due to high levels of male parental investment, humans evolved into a somewhat "androgynous" species. In this model, both males and females exhibit traits typically associated with the opposite sex. Women may display traits such as competition for mates, while men may exhibit traits like providing parental care and being selective about their mates (Stewart-Williams & Thomas, 2013).

While it is difficult to ascertain which strategy was the most prevalent during human evolution, research indicates that gender differences in mate preference are relatively stable and distinct. A cohort-longitudinal analysis of over 7,000 college students found men showed higher levels of sociosexuality and sexual permissiveness (i.e., the desire to have more sexual partners) compared to women (Sprecher et al., 2013). Although using the mean to estimate the number of desired partners tends to inflate the average, research consistently shows a trend for higher sociosexuality levels in men (for a review see Archer, 2019). According to a meta-analysis with over 1,000,000 participants across 87 countries, men consistently showed more permissive behavior towards masturbation, pornography use, casual sex, and attitudes toward casual sex (Petersen & Hyde, 2010).

During evolution, humans used both pair-bond and short-term mating strategies (Buss & Schmitt, 1993). Strategic pluralism theory posits that women's sociosexuality should be more sensitive to environmental demands and stressors than men (Gangestad & Simpson, 2000). Women are more vulnerable to the physical and social costs of short-term mating, such as the risk of sexually transmitted infections, potential loss of social status, and the need for support and resources during child-rearing. These factors contribute to a more cautious approach towards mating. Thus, in environments where biparental care is critical for offspring survival, women are likely to adopt more sociosexually restricted strategies to secure long-term partnerships that provide stability and resources for their children (Gangestad & Simpson, 2000). This adaptive behavior ensures that their offspring receive the necessary care and protection to thrive, aligning with the principles of strategic pluralism theory. A large cross-cultural study investigating over 14,000 participants from 48 nations found support for the strategic pluralism theory. Greater relational and reproductive freedom for women is associated with more moderate sex differences in sociosexuality, primarily due to women's increased promiscuity (Schmitt, 2005). Men typically do not suffer as much from social stigma for having more sexual partners. This sexual double-standard still exists today (for a review see Sagebin Bordini & Sperb, 2013), although recent research has indicated that people's attitudes towards short-term sexual strategies have shifted over the past decades in Western societies, with a general trend towards more permissive attitudes and behaviors (Twenge et al., 2015). During the 19th and 20th centuries, premarital intercourse was traditionally frowned upon, particularly for women. Besides paternal certainty, this is a key aspect of the sexual double standard, which established that premarital sex was wrong for women but allowable or right for men. Women who were sexually active before marriage were stigmatized, while men did not face the same social condemnation (Sagebin Bordini &

Sperb, 2013). Hence, the most successful strategies for producing the highest number of offspring were available to men due to their ability to have children with numerous partners. Studies on human societies confirm that polygyny has been a common marital arrangement in our evolutionary history. Comprehensive surveys of world cultures have revealed that a significant majority of pre-industrial societies practiced some form of polygyny (Murdock, 1981). Research examining ecological and social factors contributing to the prevalence of polygyny found that approximately 80% of societies in the ethnographic record have been identified as practicing polygyny (White & Burton, 1988). An evolutionary perspective also supports this, arguing that polygyny has been a common strategy for maximizing reproductive success in many human societies (Betzig, 1986). Since there is limitation (i.e., not every man can mate with an attractive woman and not every woman can mate with a high-status man), naturally, more desirable men (i.e., men of high status) will mate with more desirable women (i.e., attractive women). Historical examples illustrate this pattern. For instance, the Moroccan emperor Moulay Ismail the Bloodthirsty maintained a harem of 500 women and fathered 888 children, demonstrating how high-status men could monopolize reproductive opportunities with numerous attractive partners (Betzig, 1993). Similarly, in imperial China, emperors and high-ranking officials had access to large harems, with the number of women directly correlating with the man's status. Emperors could have thousands of women, while upper-class men had significantly fewer, highlighting the link between status and access to desirable women (Betzig, 1993).

In support of the beauty-status exchange hypothesis, it was found that desired traits for long-term partners align with gender differences. While desired traits for a long-term partner are strikingly similar between the sexes, two main differences are observed: men value attributes related to physical appeal and sexual desirability more than women, and women tend to emphasize characteristics pertaining to social status and resources (Regan et al., 2000). Ha et al. (2012) further confirmed that heterosexual men valued attractiveness of potential partners the most, while heterosexual women valued social status the most. However, there is also counterevidence for the beauty-exchange hypothesis. Data from over a thousand married, cohabiting or dating couples was analyzed to test the beauty-status exchange hypothesis. While initial observations hinted at an exchange of desired traits, a refined multivariate regression model showed that within-individual correlation of desirable traits serves as a superior predictor. In other words, when selecting mates, people tend to pay more attention to individual desirable traits such as being both physically attractive and having high SES rather than exchanging physical attractiveness for SES. Although there is evidence for the beauty-status exchange, it lacks robustness in alternative model specifications (McClintock, 2014).

It is also crucial to acknowledge the role of female choice and the limitations of mate access in sexual selection. Women invest more resources into raising their offspring, thus, having a reliable partner who is both attractive and possesses moderate SES may offer the best combination of genetic quality and commitment to ensure the survival and success of their offspring. The concept of serial pair-bonding as a reproductive strategy likely developed to address the increased reproductive burden on females, emerging after the appearance of homo sapiens to ensure infant survival through weaning. This indicates that evolutionary pressures related to reproduction and offspring survival have significantly shaped human relationships and family structures (Fisher, 1989). Particularly, highly attractive women express higher standards for traits related to good genes, good investment abilities, good parenting abilities, and good partner qualities in a long-term mate (Buss & Shackelford, 2008). This finding challenges the trade-off model, which proposes that women must choose between good genes and good investment in a long-term mate, at least for highly attractive women. Also, several factors can cause a discrepancy between preferences and actual partner choices. Generally, only those deemed most attractive are able to attract similarly attractive partners (Feingold, 1988; Luo & Zhang, 2009). Assuming that during human evolution, only a minority of men were able to attain high-status, most of the group members tended to have limited access to desirable partners. The complexity of human mating strategies, including the use of both short-term and long-term mating strategies, further complicates this issue. Therefore, while polygynous societies should have theoretically enhanced the genetic correlation between IQ and physical attractiveness, the actual historical prevalence and impact of such mating systems remain difficult to ascertain. If only a small fraction of men exclusively mated with the most physically attractive women, it is questionable how significant this effect would truly be.

"Intelligence is heritable, such that sons and daughters of more intelligent men are more intelligent than sons and daughters of less intelligent men."

Numerous studies have examined the relative contributions of genes and the environment to IQ variations, consistently pointing towards the impact of genetic factors (see "The Minnesota Twin Study" by Bouchard et al., 1990; Plomin & DeFries, 1980; Bartels et al., 2002; Sniekers et al., 2017; Hilger et al., 2022). Besides the genetic heritability component (a or h^2), IQ variance can also be explained by shared or common environmental factors (c) and environmental factors that differentiate family members from each other (e). These three components collectively form the ACE model of behavioral genetics, which is used to study the heritability of traits (Eaves et al., 1978). The heritability of IQ is relatively stable, with the genetic component being less prominent in early childhood when brain plasticity is at its peak (accounting for less than 50% of IQ variance) (Haworth et al., 2010). Throughout an individual's lifespan, the effect of the genetic component becomes larger, explaining about 60-80% of IQ heritability (Haworth et al., 2010; Plomin et al., 2014). Recent research from a meta-analysis investigating over 700,000 monozygotic-dizygotic twins suggests an average heritability of 56% across specific cognitive abilities, which is approximately similar to the heritability of *g* (Procopio et al., 2022).

Addressing the potential impact of ethnicity on IQ heritability, a recent meta-analysis investigated the heritability of IQ between different ethnic groups (Whites, Blacks and Hispanics living in the United States) and found a moderate to high heritability of IQ, which did not significantly differ between ethnic groups (Pesta et al., 2020).

"Beauty is heritable, such that sons and daughters of more beautiful women are more beautiful than sons and daughters of less beautiful women."

Studies investigating the heritability of physical attractiveness of monozygotic and dizygotic twins report high correlations, ranging from $r = 0.54$ (McGovern et al., 1996) to $r = 0.94$ (Rowe et al., 1989). Recent research indicates that the heritability of specific facial traits range between 28% to 67% (Cole et al., 2017). Evidence from a large-scale study of over 900 twins showed that heritability estimates reached or even surpassed 0.7 for specific regions such as the chin area, nasal regions, nasolabial folds, upper lips, and zygomatic bones (Tsagkrasoulis et al., 2017). Purkey (2011) investigated a sample of over 1,500 twins and found that facial attractiveness was best explained by the AE model (genetic component x environmental factors), with the genetic component accounting for 65% of the variation. Notably, no significant sex differences in the heritability of facial attractiveness were observed. A genome-wide association study (GWAS) identified significant associations between facial attractiveness and genetic loci including FC-AS (6p25.1) and MC-AS (20q13.11 and 2q22.1). Additionally, two genome-wide significant loci associated with facial attractiveness were identified on 10q11.22 and 2p22.2. Sex-specific associations with various social factors were revealed, suggesting that different genetic components may be associated with male and female attractiveness. This notion gains further support from the work of Hu et al. (2019), which shows a sex-specific genetic architecture of facial attractiveness. Overall, while the exact heritability of facial attractiveness may vary depending on the population and methodology used, the available evidence suggests that genetic factors may contribute to determining facial attractiveness.

Evidence from prior Meta-Analyses on the Attractiveness-Intelligence Association

Previous meta-analyses, such as those conducted by Eagly et al. (1991), Feingold (1992) and Langlois et al. (2000) explored the connection between attractiveness and various metrics of intelligence, including perceived or self-rated attractiveness as well as proxies such as grade-point average (GPA), job performance and standardized test scores like the SAT. However, these analyses encountered challenges due to the inclusion of heterogeneous measures of intelligence, leading to potential distortion of the results. Recognizing the need for a more nuanced approach, the present meta-analysis focuses exclusively on studies using psychometric IQ-tests to assess intelligence. By narrowing the scope in this manner, the aim is not only to refine the investigation but also to address the issue of comparability across studies. Importantly, this meta-analysis seeks to explore the stability of the attractiveness-intelligence relationship across various moderators such as age, gender and attractiveness assessment, employing multiverse analysis to assess the robustness of findings.

Methods

The present meta-analysis was preregistered prior to accessing the data. The pre-registration protocol is available at the Open Science Framework (OSF; https://osf.io/2asdn/). Any deviations from the preregistration protocol are listed in Appendix B. Study quality was assessed using the Newcastle-Ottawa Scale (NOS; Herzog et al., 2013), which is accessible in Appendix D, Table S2. Following hypotheses were formulated:

H1: *There is a positive association between facial attractiveness and IQ.*

H2: *There is a positive association between childhood attractiveness and adult IQ.*

H3: *There will be a difference in the association between attractiveness and IQ across different age groups.*

H4: *There will be a difference in the association between attractiveness and IQ based on participants' sex.*

Previously included factors for subgroup analyses stated in the preregistration protocol (see OSF; https://osf.io/2asdn/) (i.e., moderator effects for sex, intelligence tests, and status of publication) were also not analyzed due to missing or insufficient data.

Literature search

Potentially relevant studies were searched in three databases (PubMed, Scopus, ISI Web of Science) as well as the Open Access Dissertation and Theses (https://oadt.org) for grey literature. The following search string was used for the identification of potentially eligible articles (intelligen* AND beaut*) OR (intelligen* AND attractive*) OR (IQ AND beaut*) OR (IQ AND attractive*). Additionally, a cited reference search of the article by Kanazawa & Kovar (2004) was conducted. Titles and abstracts of 1,994 potentially relevant articles were screened and full-text versions of 91 studies were obtained (for a flowchart, see Fig. 1; excluded records are listed in Appendix E, Table S3). The literature search was performed in May 2023 and updated in July 2023. Another update was conducted in September 2023.

Inclusion criteria

Studies had to meet five inclusion criteria in order to be eligible for inclusion in the present meta-analysis. First, studies had to investigate the association between intelligence and facial attractiveness using Pearson correlation *r*. Second, intelligence had to be measured with psychometric tests (i.e., IQ-tests). Studies that used other methods to measure intelligence (e.g., academic achievement, job status, intellectual competence, grade point average etc.) were excluded. Thirdly, attractiveness had to be rated in terms of facial attractiveness. Ratings of facial attractiveness had to be either dichotomous or polytomous. Live ratings of attractiveness (i.e., face to face ratings), photographs and videos were included. Fourthly, studies that measured body attractiveness (e.g., hip-to-waist ratio, shoulder width, etc.), physical grooming or fashion style were also excluded. There were no age, sex or population restrictions. Both adult and child samples were included. Finally, sufficient statistical information to calculate the effect size had to be available.

Coding

Coding was conducted twice independently by a trained researcher [P.T.]. Information was coded for the included studies: (i) study characteristics (publication status: published vs. unpublished; publication year; manuscript type: journal vs. thesis vs. others; peer-reviewed: yes vs. no; funding: yes vs. not reported;), (ii) country, (iii) sample descriptors (sample size; mean age; sample type: children vs. adults vs. mixed; percentage of males within sample;), (iv) intelligence measurements (IQ-test; mean; standard deviation; reliability;), (v) attractiveness measurements (attractiveness type: live vs. photographs vs. videos;

attractiveness ratings: dichotomous vs. polytomous; raters; attractiveness scale; reliability;), (vi) statistical parameters (effect size; *p*-values;). In case of missing information, corresponding authors were contacted twice (two weeks after the first contact) if authors did not respond. If the authors did not respond or if there was no active email address was reported, respective studies were excluded. The coding file is available on OSF (https://osf.io/2asdn/).

Data analysis

Analysis was conducted using the free open-source software R 4.3.1 (R Core Team, 2022). First, effect sizes were transformed into Fisher's *z* and synthesized by precision-weighted random effects model. I^2 was calculated as a descriptive measure of heterogeneity and interpreted for 25%, 50%, and 75% representing lower thresholds of small, moderate or large heterogeneity, respectively (Higgins et al., 2003) and prediction intervals (Borenstein, 2021). Afterwards, *z*-transformed coefficients were transformed back to the *r*-metric to facilitate interpretation.

Subgroup analyses were conducted to assess the influence of moderator variables using a mixed-effects model (based on random-effects models). Effect sizes of the subgroup analyses were grouped according to sample type (children vs. adults vs. mixed) and attractiveness assessments (live vs. photographs) (see Table 3). For the moderator analysis of the categorical variable "sample type" (children vs. adults vs. mixed), "adults" was used as the reference group, while the other two categories were dummy-coded (see Table 3). Recent evidence in psychological literature suggests a decline-effect over time (Schooler, 2011; Pietschnig et al., 2019). Publication year as a continuous moderator was explored to ascertain the presence of a decline-effect (see Table 4). Additionally, mean age was examined through meta-regression to determine its impact as a moderator on the relationship between intelligence and attractiveness (see Table 4). The possible impacts of leverage points were evaluated using leave-one-out analyses (see Table 5).

To detect dissemination bias, several bias detection methods were performed. Since the final sample only included published studies, the full dataset was analyzed for publication bias. First, funnel plots were generated to visually detect indications of asymmetry. Egger's regression test (Sterne & Egger, 2005) and the rank correlation method (Begg & Mazumdar, 1994) were used to examine potential funnel plot asymmetries. The trim-and-fill method (Duval & Tweedie, 2000) was deployed to estimate the number of 'missing' studies in an asymmetric funnel plot.

Furthermore, three methods to assess publication bias (*p*-curve, *p*-uniform, *p*-uniform*) were used to estimate meta-analytical summary effects based on *p*-value distributions of significant published studies (Simonsohn et al., 2014; van Assen et al., 2015; van Aert & van Assen, 2021). The *p*-curve method only uses significant *p*-values ($p < .05$) to create a distribution of significant *p*-values across a set of studies to show whether there is evidence for a true effect. It estimates effect sizes by aligning the theoretical distribution with the observed *p*-values. In presence of publication bias, the average true effect size deviates from the average observed effect size (Simonsohn et al., 2014). Right-skewed *p*-curves indicate that the studies in the set yielded significant results, implying that a true effect is detectable. Left-skewed curves reveal potential *p*-hacking or questionable research practices. If the *p*-values are uniformally distributed across the range, it suggests a lack of evidence. In other words, the studies in the set do not provide meaningful information to support or reject the hypothesis. The approach is also used to detect potential *p*-hacking, indicated by a reduced right skew of the *p*-curve (Simonsohn et al., 2014). To calculate *p*-curve, I used the *p*-curve app (Simonsohn et al., 2015; https://p-curve.com).

P-uniform corrects for publication bias in meta-analyses by leveraging the distribution of *p*-values from the included studies. It evaluates the distribution of conditional *p*-values to detect any left-skewness, which can indicate publication bias, within a fixed-effect model framework. The method assumes that, under the null hypothesis, the *p*-values should form a uniform distribution. By using this property, *p*-uniform estimates the effect size while adjusting for potential bias introduces by selective reporting (van Assen et al., 2015; McShane et al., 2016).

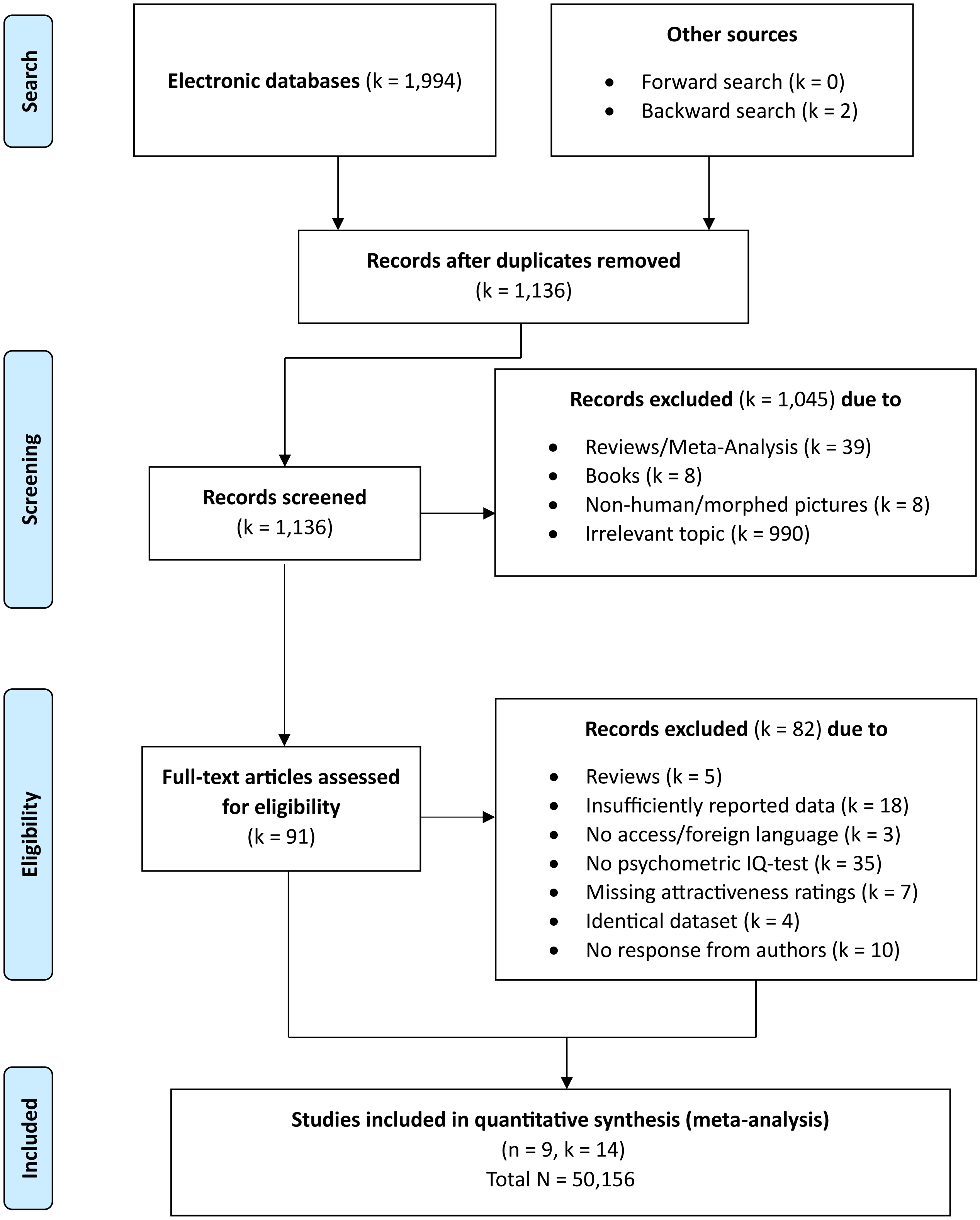
P-uniform* works in a similar manner as *p*-uniform, but additionally accounts for heterogeneity in the data, making it more flexible and applicable to a wider range of meta-analyses. It incorporates both significant and non-significant studies, enhancing accuracy and reducing overestimation due to heterogeneity. *P*-uniform* corrects for publication bias by calculating the effect size under the assumption that, under the null hypothesis, the *p*-values from the studies should be uniformly distributed. This correction enables a more precise estimation of the actual effect size, even when there is heterogeneity among the studies (van Aert & van Assen, 2021) (for a detailed description of publication bias detection methods see Table S4, Appendix G).

Specification curve analysis (SCA) was employed to test different combinations of how the data was analyzed (internal *"which"* and external *"how"* factors) (see https://osf.io/nkv46/ for the R Code from Voracek et al., 2019). These specifications are characterized by being sensible tests of the research question, expected to be statistically valid and non-redundant with others in the set. The graphic results display the estimated effect sizes across all specifications, sorted by magnitude, with a corresponding “dashboard chart” elucidating the operationalizations behind each result. This visual representation facilitates the identification of influential studies. The test evaluates whether, collectively, the specifications reject the notion that the effect is non-existent. In other words, SCA shows whether the total number of results allow the rejection of the null hypothesis. This also allows the comparison of results across different specifications. SCA comprises of three steps, (i) the identification of reasonable specifications (which data to analyze, and how), (ii), visualization and description of the results, and (iii) inference of conclusion across all specifications (Simonsohn et al., 2020). In the present meta-analysis, only two internal factors were analyzed: i) sample type (children vs. adults vs. mixed), and ii) attractiveness assessment (live vs. photographs). These specifications yielded 3 * 2 = 6 combinations. The external factors were: i) effect size (i.e., Pearson’s *r* and *z*-transformed coefficients), and ii) estimator type (i.e., random-effects DerSimonian-Laird vs. random-effects restricted maximum-likelihood estimators vs. fixed-effect estimation vs. unweighted estimation). These specifications resulted in 2 * 4 = 8 combinations to analyze the same data. In total, the combination of factors yielded 6 * 8 = 48 unique ways to analyze the data.

Combinatorial meta-analysis was used to investigate between-study heterogeneity and to identify potential influential studies. In this approach, separate meta-analyses are performed on all possible subsets of studies that are included in the meta-analysis. Graphical Display of Study Heterogeneity plots (GOSH-plots) are used as visualizations for the possible subsets. The method involves reporting the results for all or a significant random subset of “reasonable specifications” $2^k - 1$ (Olkin et al., 2012). In the present study, I obtained (i) $2^{14} - 1 = 16{,}383$ combinations for the full dataset, and (ii) $2^{13} - 1 = 8{,}191$ for the trimmed dataset. GOSH-plots were used to visually explore the heterogeneity among studies and to detect any potential subpopulations with differing average effect sizes. By examining the distribution of effect sizes across all possible subsets of studies, GOSH-plots can reveal patterns that might not be apparent from a single meta-analysis. For instance, the presence of two or more distinct clusters within the plot may suggest the existence of subpopulations that vary in their average effect size (Olkin et al., 2012). This visual inspection helps in understanding the underlying structure of the data and in identifying studies that may disproportionately influence the overall results.

One study (Kanazawa, 2011 – Group A) presented an exceptionally high correlation coefficient (*r* = .38), contrasting with the consistently lower correlations found in other studies encompassed within this meta-analysis. To assess the impact of this leverage point, the data was analyzed twice: initially including all studies (*k* = 14), and subsequently with a trimmed dataset excluding the outlier study (*k* = 13). This approach aims to ensure the robustness of the analysis by examining the potential effects of this leverage point.

Figure 1. *Flow-chart of study inclusion.*



Final sample

The final sample included a total of nine studies with 14 subsets (N = 50,156). Studies that were excluded due to various reasons can be found in Table 3, Appendix E. In total, the present meta-analysis contains five children-only samples, eight adults-only samples and one mixed-age sample. The age of participants ranged from seven to 60 and the majority of participants were from the United States of America (N = 30,489). Only three studies (k = 3) used live ratings of attractiveness, the remaining studies used photographs (k = 11) depicting participants' faces with neutral expression under standardized conditions. While most studies deployed a Likert-Scale to rate attractiveness (k = 13), only one study used a forced-choice, dichotomous rating (k = 1) conducted by two elementary teachers of the students (Kanazawa, 2011 – Group A). A detailed breakdown of attractiveness assessment methods can be found in Table 1. While in most included studies, fellow colleagues or college undergraduates rated attractiveness (k = 8), one study used "trained coders" (Härtel et al., 2023). Regarding the assessment of intelligence, the most common used IQ-tests were the Stanford-Binet Intelligence Scale, the Wechsler Adult Intelligence Scale (WAIS-III, WAIS, WAIS-R), the Raven Progressive Matrices Test (RPM), the Multidimensional Aptitude Battery (MAB), Intelligence Structure Test 2000 R (IST-R 2000) and the Henmon-Nelson Test of Mental Ability (see Table 1 for a detailed description of applied IQ-tests, for further description see Appendix F).

Table 1. *Details of included studies.*

Reference	*N*	Sample type	*r*	*p*-value	Intelligence Measurement	Attractiveness Measurement	Attractiveness Judges	% of Males	Journal
Kanazawa (2011) – Group A	17,419	Children	.381	< .001	Age 7: Copying Designs Test, Draw-a-Man Test, Southgate Group Reading Test, Problem Arithmetic Test Age 11: Verbal General Ability Test, Nonverbal General Ability Test, Reading Comprehension Test, Mathematics Comprehension Test, Copying Designs Test Age 16: Reading Comprehension Test, Mathematics Comprehension Test	live	Two elementary school teachers	n/a	Intelligence
Kanazawa (2011) – Group B	20,745	Mixed	.126	< .001	Peabody Picture Vocabulary Test (PPVT)	live	Interviewer	n/a	Intelligence
Colladon et al. (2018)	200	Adults	-.042	n/p	Raven's Advanced Progressive Matrices test	n/a	Colleagues/fellow students	51%	International Journal of Entrepreneurs hip and Small Business
Lee et al. (2017)	1,660	Children	.01	.517	The Multidimensional Aptitude Battery (MAB)	photographs	Seven or 14 raters	n/a	Intelligence
Judge et al. (2009)	191	Adults	.16	< .05	Wechsler Adult Intelligence Scale (WAIS) Schaie-Thurstone Letter Series test Raven's Advanced Progressive Matrices test	photographs	Six undergraduates	n/a	Journal of Applied Psychology
Kleisner et al. (2014)	80	Adults	-.003	n/p	Intelligence Structure Test 2000 R (ISTR-2000)	photographs	160 (75 men, 85 women) raters	50%	PloS one

Zebrowitz et al. (2002)	N_1 = 186 N_2 = 176 N_3 = 170 N_4 = 123	Mixed	.26 .16 .21 .22	< .001 ≤ .05 ≤ .01 ≤ .01	Childhood, puberty, adolescence: Stanford-Binet Adulthood: Wechsler Adult Intelligence full scale (WAIS-R)	photographs	College undergraduates	46.24% 47.73% 47.06% 54.53%	Personality and Social Psychology Bulletin
Gangestad et al. (2010)	66	Adults	.18	n/p	Raven’s Advanced Progressive Matrices, 12-Item Version	photographs	20 (10 male, 10 female) raters	100%	Evolution and Human Behavior
Härtel et al. (2023)	311	Adults	-.02	n/p	Raven's Advanced Progressive Matrices, 15-Item Version German multiple-choice vocabulary test B (Mehrfachwahl-Wortschatz-Test B, MWTB)	photographs	Six trained coders	47.92%	European Journal of Personality
Scholz & Sicinski (2015)	8,623	Adults	.031	< .1	Henmon-Nelson Test of Mental Ability	photographs	12 (six male, six female) raters	100%	The Review of Economics and Statistics

Note. N = sample size; r = Pearson correlation coefficient;

Results

Main analyses

Random-effects analyses were calculated for the full data and for the trimmed dataset (without the dataset from Kanazawa (2011) – Group A since it yielded an exceptionally high correlation between intelligence and attractiveness (r = .38)) (see Fig. 2A, 2B, Table 2). The full dataset condition resulted in a significant summary effect of r = .13 (k = 14, p = .0005, 95% CI [0.06, 0.20]), showing a small correlation between intelligence and attractiveness. Between-study heterogeneity was substantial for the full dataset (k = 14, Q = 1155.20, p < .0001, I^2 = 97.53%, τ^2 = 0.012, SE = 0.008). Outlier analysis by the leave-one-out method revealed the influential impact of the dataset by Kanazawa (2011) – Group A on overall results (see Table 5). Analysis excluding this outlier also showed a significant summary effect of r = .10 (k = 13, p = .0004, 95% CI [0.04, 0.15]), albeit more modest in magnitude. Between-study heterogeneity of the trimmed dataset was greatly reduced (k = 13, Q = 88.32, p < .0001, I^2 = 88.96%, τ^2 = 0.006, SE = 0.004). The potential moderating effects of age (children vs. adults vs. mixed) and attractiveness assessment (live vs. photographs) on the correlation between attractiveness and intelligence were analyzed for both the full dataset (k = 14) and the trimmed dataset (k = 13) (see Table 3). Overall and subgroup summary effects are provided in Table 2.

Subgroup analyses

Results showed that the effect was stronger in children r = .21 (p = .003, 95% CI [0.07, 0.35]) compared to adults r = .06 (p = .047, 95% CI [0.0008, 0.12], see Table 2), indicating that the relationship between intelligence and attractiveness is more pronounced in children. The heterogeneity was moderate in adult samples (I^2 = 47.23%), suggesting some variability in the effect sizes across studies, whereas it was very high in children samples (I^2 = 95.70%), indicating substantial variability.

As for the method of attractiveness assessment, live ratings yielded a higher correlation r = .17 (p = .188, 95% CI [-0.08, 0.42]) compared to ratings based on photographs r = .11 (p = .001, 95% CI [0.04, 0.17], see Table 2). However, the results for live ratings were insignificant (p = .188). This suggests that the observed higher correlation for live ratings may be due to random variation rather than a true effect. In contrast, the effect for photograph ratings was significant, suggesting that there is a true, albeit modest, association between intelligence and attractiveness when attractiveness is assessed via photographs.

Moderator analysis

Moderator analysis showed that the sample type (children vs. adults vs. mixed; adults used as reference category) explained only 11.83% of variance in the relationship between intelligence and attractiveness (see Table 3). There was a marginal difference between the effect of children and adults, with children-only samples tending to show a stronger effect r = .14 (p = .064, 95% CI [-0.008, 0.29]), although this finding was not statistically significant. No nominally significant influence was observed between mixed and adult samples r = .07 (p = .141, 95% CI [-0.02, 0.17]). Overall, while the sample type shows a marginally stronger effect for children compared to adults, the influence of sample type on the relationship between IQ and attractiveness is not nominally significant (Q = 3.45, df = 2, p = .181). Similarly, although the overall effect for live ratings was positive r = .18 (p = .015, 95% CI [0.04, 0.33], see Table 3), the type of attractiveness assessment (live vs. photographs) did not have a statistically significant influence on the relationship between intelligence and attractiveness (Q = 0.613, df = 1, p = .434). It is important to note that among the three studies using live assessments, the outlier study by Kanazawa (2011 – Group A) was included, which may have skewed the results. The type of attractiveness method also only explained 0.64% of variance in the association between intelligence and attractiveness (see Table 3).

Initially, several other potential moderators, including intelligence assessment methods (e.g., Stanford-Binet vs. WAIS, etc.), gender (men only vs. women only), and publication status (published vs. unpublished) were intended for investigation. However, this proved unfeasible due to missing and/or insufficient data (see Appendix B).

Publication year as a continuous moderator was examined by linear precision-weighted meta-regression and yielded a significant influence on the association between attractiveness and intelligence (b = -.012, Q = 885.79, R^2 = 26.71%, p = .018, see Table 4). In other words, as the years of publication increase, the effect between attractiveness and intelligence seems to decrease. Mean age as a continuous moderator did not yield a significant influence on the association between attractiveness and intelligence (b = -.002, Q = 501.50, R^2 < 0.01%, p = .524, see Table 4). The investigation into the male ratio was initially intended. However, the analysis of this aspect was deferred due to insufficient data availability (see Appendix B).

Figure 2. *Forest plots. Panel A displays the full dataset (k = 14), panel B shows the trimmed dataset without the outlier dataset (k = 13).*

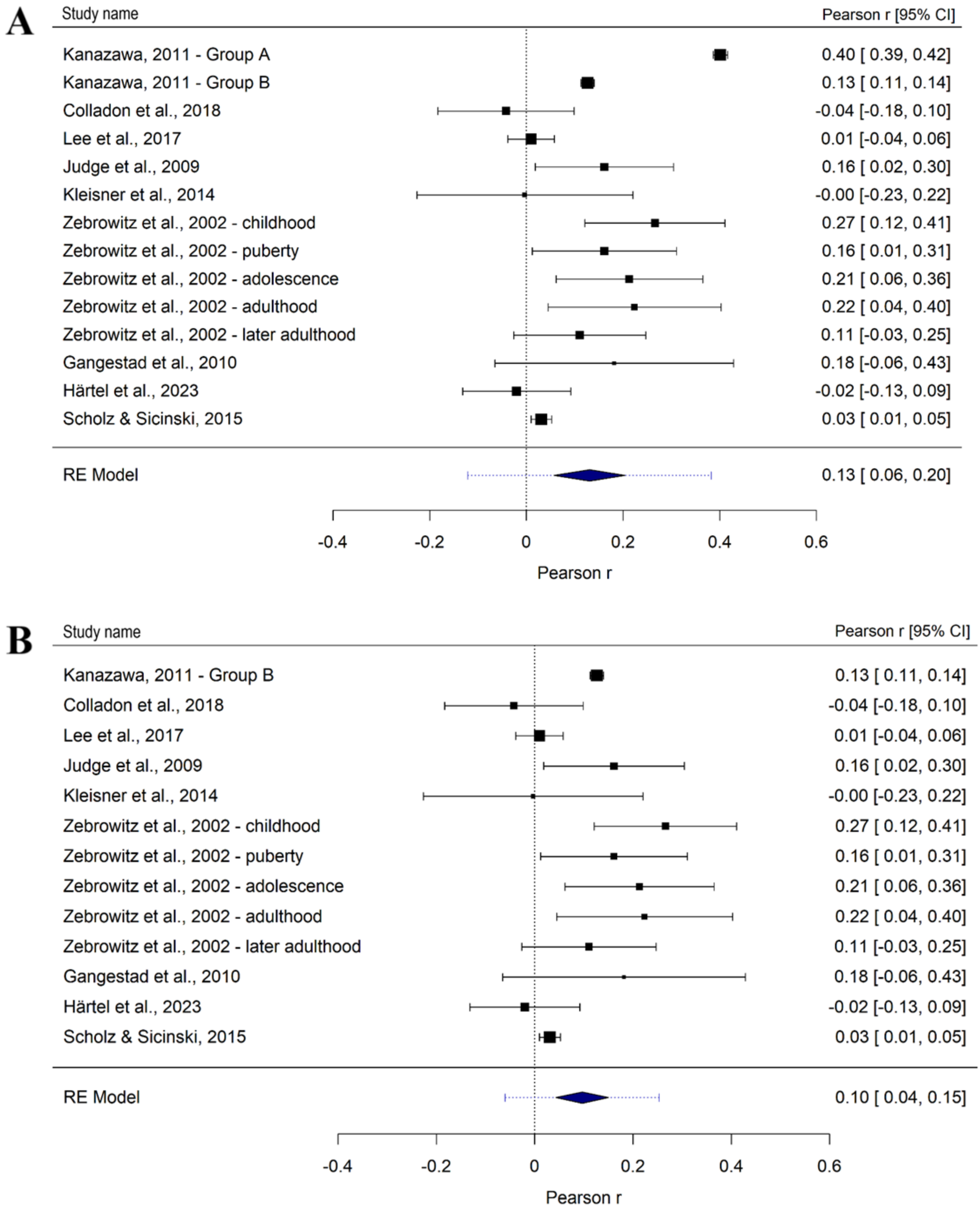


Table 2. *Random-effects estimates of overall data and subgroup analyses.*

	Summary effect (*r*)	***SD***	***p*-values**	**95% CI**	***Q***	***τ² (SE)***	***I²***
Full dataset (k = 14)	.131	0.037	.0005	[0.06, 0.21]	1155.204***	0.0151 (0.008)	97.53%
Trimmed dataset (k = 13)	.097	0.273	.0004	[0.04, 0.15]	88.324***	0.0056 (0.004)	88.96%
Adults (k = 8)	.061	0.031	.047	[0.0008, 0.12]	12.138*	0.0029 (0.004)	47.23%
Children (k = 5)	.211	0.072	.003	[0.07, 0.35]	244.428***	0.0224 (0.181)	95.70%
Mixed (k = 1)	.217	0.007	< .0001	[0.11, 0.14]	0.000	0.000 (0.000)	0%
live (k = 3)	.169	0.128	.188	[-0.08, 0.42]	730.317***	0.0476 (0.049)	99.78%
photographs (k = 11)	.106	0.032	.001	[0.04, 0.17]	29.453***	0.0065 (0.005)	77.16%

Note. SE = standard error; 95% CI = 95% lower and upper bound of 95% confidence interval; Q = Cochran's Q test statistic for heterogeneity; τ^2 = between-studies variance; I^2 = ratio between true heterogeneity and observed variation;

Table 3. *Numeric outcomes of moderator analyses.*

Predictor	*r*	R^2	*p*-value	95% CI	*Q*	*QM*	*τ (SE)*	I^2
Age (reference group: adults)	.072	11.83%	0.141	[-0.02, 0.17]	256.566***	3.425 (*df* = 2)	0.115 (0.008)	88.12%
Children	.139		0.064	[-0.008, 0.27]				
Mixed	.055		0.664	[-0.19, 0.30]				
Attractiveness assessment (live)	.181	0.64%	0.015*	[0.04, 0.33]	759.770***	.613 (*df* = 1)	0.123 (0.008)	96.83%
Photographs	-0.07		0.434	[-0.24, 0.10]				

Note. b = unstandardized regression coefficient; R^2 = proportion of variance in the dependent variable explained by the independent variable; *SE* = standard error; 95% CI = 95% lower and upper bound of 95% confidence interval; Q = Cochran's Q test statistic for heterogeneity; *QM* = test of moderators; τ^2 = between-studies variance; I^2 = ratio between true heterogeneity and total observed variation; *p < .05; ***p < .001.

Table 4. *Numeric outcomes of linear-precision-weighted meta-regression.*

Predictor	*b* (SE)	R^2	*p*-value	95% CI	*Q*	τ (SE)	I^2
Publication year	-0.012 (0.005)	26.71%	.018*	[-0.02, -0.002]	885.785***	0.011 (0.006)	96.41%
Age (mean)	-0.002 (0.003)	< 0.01%	.524	[-0.008, 0.004]	501.497***	0.016 (0.008)	96.38%

Note. *b* = unstandardized regression coefficient; R^2 = proportion of variance in the dependent variable explained by the independent variable; SE = standard error; 95% CI = 95% lower and upper bound of 95% confidence interval; *Q* = Cochran's *Q* test statistic for heterogeneity; τ^2 = between-studies variance; I^2 = ratio between true heterogeneity and total observed variation;

Table 5. *Leave-one-out analyses.*

Study name	Summary effect	*SE*	*p*-value	95% CI	*Q*	τ^2	I^2
Kanazawa (2011) – Group A	r = 0.097	0.027	.0004	[0.04, 0.15]	88.324***	0.006	88.96%
Kanazawa (2011) – Group B	r = 0.130	0.041	.0013	[0.05, 0.21]	958.086***	0.017	95.81%
Colladon et al. (2018)	r = 0.142	0.038	.0002	[0.07, 0.22]	1143.672***	0.014	97.60%
Lee et al. (2017)	r = 0.141	0.039	.0003	[0.07, 0.22]	1092.488***	0.015	97.55%
Judge et al. (2009)	r = 0.128	0.040	.0013	[0.05, 0.21]	1154.903***	0.016	97.86%
Kleisner et al. (2014)	r = 0.137	0.039	.0004	[0.06, 0.21]	1154.985***	0.015	97.75%
Zebrowitz et al. (2002) – childhood	r = 0.120	0.039	.0018	[0.05, 0.20]	1151.432***	0.015	97.71%
Zebrowitz et al. (2002) - puberty	r = 0.128	0.040	.0013	[0.05, 0.21]	1154.927***	0.016	97.86%
Zebrowitz et al. (2002) – adolescence	r = 0.124	0.039	.0016	[0.05, 0.20]	1154.180***	0.016	97.81%
Zebrowitz et al. (2002) – adulthood	r = 0.125	0.039	.0015	[0.05, 0.20]	1155.143***	0.016	97.80%
Zebrowitz et al. (2002) – later adulthood	r = 0.131	0.040	.0009	[0.05, 0.21]	1155.496***	0.016	97.87%
Gangestad et al. (2010)	r = 0.127	0.039	.0010	[0.50, 0.20]	1153.180***	0.016	97.81%
Härtel et al. (2023)	r = 0.142	0.038	.0002	[0.07, 0.22]	1140.025***	0.015	97.61%
Scholz & Sicinski (2015)	r = 0.140	0.040	.0004	[0.06, 0.22]	853.288***	0.016	96.85%

Note. SE = standard error; 95% CI = 95% lower and upper bound of 95% confidence interval; *Q* = Cochran's *Q* test statistic for heterogeneity; τ^2 = between-studies variance; I^2 = ratio between true heterogeneity and observed variation;

Dissemination Bias

The full dataset was analyzed for publication bias (k = 14) since all studies included were previously published. Out of the seven methods that were used to assess dissemination bias, most did not show strong indications of publication bias. The presence of the leverage study (Kanazawa, 2011 – Group A, r = .38) may have masked some indications for publication bias, and inflated the overall effect size estimate. However, evidence for publication bias was also not substantial in the trimmed analysis. Numerical outcomes of publication bias detection methods are provided in Table S4, see Appendix G.

While the funnel plot shows no evidence for publication bias in the full data analysis (Fig. 3A), the funnel plot becomes slightly asymmetric in the trimmed dataset, adding 3 studies on the left side of the funnel plot (Fig. 3C). This asymmetry highlights the influence of the leverage point (Kanazawa, 2011 – Group A, r = .38) on the overall results. Trim-and-fill analyses also indicate no missing studies on the right side of the funnel plot for the trimmed dataset, but the funnel plot shows slight asymmetry in the trimmed dataset, revealing 3 missing studies on the left side. This suggests that including the outlier dataset may have masked the presence of publication bias. Power-enhanced funnel plots were used to assess the average power of the included studies. Figure 3B and 3D show that most studies had lower power, particularly in the trimmed analysis (Fig. 3D). The median power of the full analysis was 43.8%, while for the trimmed analysis, it dropped to 26.3%, likely due to the exclusion of the outlier study, which had a large effect size and contributed to the overall power of the meta-analysis. The Replicability-Index was 30.5% for the full analysis, suggesting a low probability of replicating the average individual effect size. In the trimmed analysis, the Replicability-Index dropped to 0%, due to the exclusion of the leverage point. The test of excess of significance revealed that there were 8 significant findings out of 14 studies (p = .999), whereas the expected number of significant findings was 11.69 based on the estimated power (within-study average power = 86.7%). The observed-to-expected ratio was 0.684, indicating fewer significant results than anticipated. For the trimmed analysis, the within-study average power dropped to 27.04%, again highlighting the influence of the leverage point. Due to the limited power, studies had a limited ability to detect a true effect in the trimmed analysis. Paradoxically, studies of the trimmed analysis were more successful in finding significant results (7 significant findings, expected number of significant findings = 5.43, p = .208). However, both the full and the trimmed analysis showed non-significant p-values, indicating no strong evidence of excess significant findings in either analysis. Results from Egger's regression test did not reveal any asymmetry either for the full analysis (Z = -

.660, p = .522), or for the trimmed analysis (Z = -.201, p = .845). The rank correlation method, however, showed evidence for publication bias in the full analysis (τ = .427, p = .036). When excluding the leverage point, indices for publication bias diminishes (τ = .256, p = .252).

P-curve, based on 8 significant studies, indicated no evidence for publication bias and *p*-hacking. The right-skewed distribution of *p*-values (see Fig. 4), with a higher frequency of small values (i.e., those close to p = .01) compared to larger *p*-values (i.e., those close to p = .05), suggests the presence of true underlying effects rather than *p*-hacking or selective reporting. *P*-uniform also showed little evidence for *p*-hacking in the full dataset and no evidence for publication bias in the trimmed analysis. The adjusted effect size was r = .126 (p = .003) for the full analysis, and r = .116 (p = .009) for the trimmed analysis. The test for publication bias was significant in the full analysis (p = .050), indicating potential *p*-hacking. This evidence for *p*-hacking disappeared in the trimmed dataset (p = .620), underscoring that the evidence for publication bias is weaker when the outlier study is excluded. The adjusted estimate from *p*-uniform* was r = .139 for the full analysis compared to a substantially lower estimate of r = .081 for the trimmed analysis. The test of the null hypothesis was significant in the full analysis (p = .035), while it becomes marginally insignificant in the trimmed analysis (p = .053). The adjusted estimate is likely inflated by the outlier study in the full analysis.

In summary, the rank correlation method and *p*-uniform suggest potential publication bias in the full dataset, however this evidence is not consistent and robust across all methods. The inclusion of the outlier study (Kanazawa, 2011 – Group A) has obscured some indications of publication bias and inflated the overall effect size estimate. When this outlier is removed, the evidence for publication bias yields mixed results. In the trimmed dataset, methods such as the funnel plot, the trim-and-fill method, Egger's regression test, and *p*-uniform* show slight indications of bias. This emphasizes the impact of the leverage point (Kanazawa, 2011 – Group A) on the true effect size. The influence of this outlier can also be observed by the significant drop in average power as shown by the funnel plots. Additionally, the decrease in within-study average power from 86.7% in the full dataset to 27.04% in the trimmed analysis highlights the substantial effect of the outlier on the power and reliability of the findings. Overall, the results show that there are no strong indications for substantial publication bias in the present meta-analysis.

Figure 3. *Contour-enhanced funnel plots with imputed trim-and-fill values and Egger's regression line (left, Fig. 3A, 3C) and power-enhanced funnel plots (right, Fig. 3B, 3D). The top line shows funnel plots of the full dataset (k = 14), the bottom line displays the trimmed dataset (k = 13). Study power ranges from 0 to 10 percent (dark red) to 80 to 90 percent (dark green) in 10 percent increments. Median average power was 43.8% for the full analysis (Fig. 3B), and 26.3% for the trimmed analysis (Fig. 3D).*

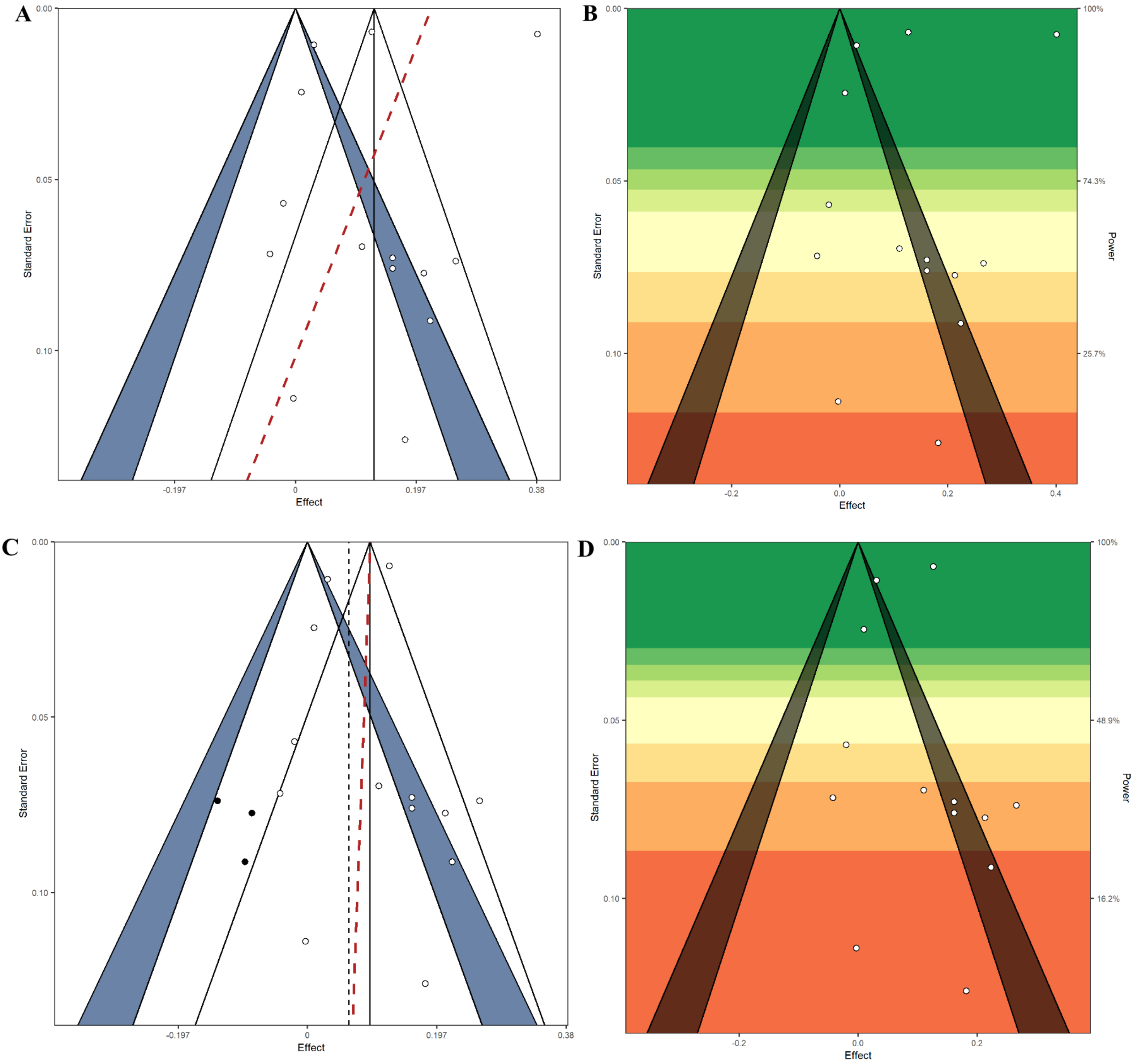
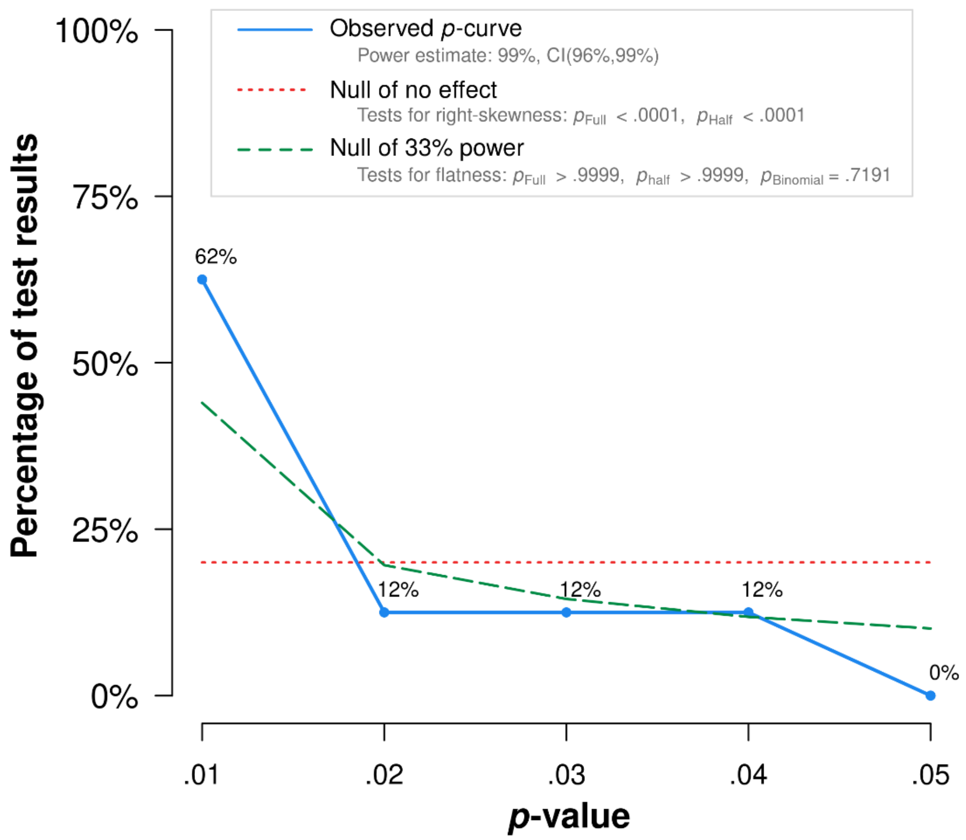


Figure 4. *p-Curve based on eight statistically significant results (p < .05), of which six are p < .025. Six additional records were excluded from p-curve because they were p > .05. The blue line represents the distribution of observed p-values. The red dashed line represents the expected distribution under the null hypothesis of no effect. The green dashed line displays the condition of a true effect and underpowered studies (33% power).*



Specification curve analysis

Specification curve analyses are displayed in Fig. 5. All 48 specifications yielded a positive summary effect (see Fig. 5A), meaning that regardless of the analytical choices made, the association between intelligence and attractiveness remained positive. None of these specifications led to a null or negative effect, indicating the robustness of the effect. This consistency suggests that the observed effect is not highly sensitive to the specific analytical choices made. Similarly, analysis of the trimmed dataset also revealed positive summary effects (see Fig. 5B). This further supports the robustness of the positive association between intelligence and attractiveness. While the analysis is robust across the specifications tested, the limited number of internal factors analyzed means that the robustness could be tested more thoroughly with additional factors. Typically, more than two internal factors (age and attractiveness assessment) are included in specification curve analyses to provide a more nuanced picture of robustness. However, the consistent positive effects across all specifications indicate a robust association between intelligence and attractiveness. Overall, the results reinforce the conclusion that more attractive people tend to be more intelligent, regardless of the specific analytical choices.

Combinatorial meta-analysis

Combinatorial meta-analyses are visualized in Figure 6 (GOSH plots). Analysis of the full dataset yielded 16,383 different combinations, a relatively small number for meta-analyses. Overall heterogeneity was substantial for the full dataset (k = 14), which suggests that the effect sizes across the included studies vary considerably. A bimodal pattern can be observed in the full dataset analysis, indicating that there are likely two distinct subgroups or clusters of effect sizes within the dataset. The reduction in heterogeneity in the trimmed dataset (k = 13), as shown by largely overlapping distributions, suggests that the leverage point was a major contributor to the observed heterogeneity. Removing this outlier has resulted in a more homogenous set of studies (Fig. 6B). This suggests that the remaining studies (k = 13) have more consistent effect sizes, leading to a more reliable and interpretable meta-analytic result. After removing the leverage point (Kanazawa, 2011 – Group A), the remaining studies form a more homogenous set, which indicates that the true effect size of the association between intelligence and attractiveness is more consistent across these studies. The reduction in heterogeneity improves the robustness of the meta-analytic conclusions. This suggests that the observed association between intelligence and attractiveness is more robust and not primarily driven by extreme values from a single study. The small spike in the trimmed dataset suggests that there may still be minor variations or subgroups within the remaining studies. However, these variations are less pronounced than in the full dataset, indicating that the overall effect size is more stable.

Figure 5. *Specification Curve Analysis. The bottom panel represents "which" and "how" factors. The middle panel represents the number of samples within the subsets. The top panel displays the estimated summary effects with respective 95% of confidence intervals.*

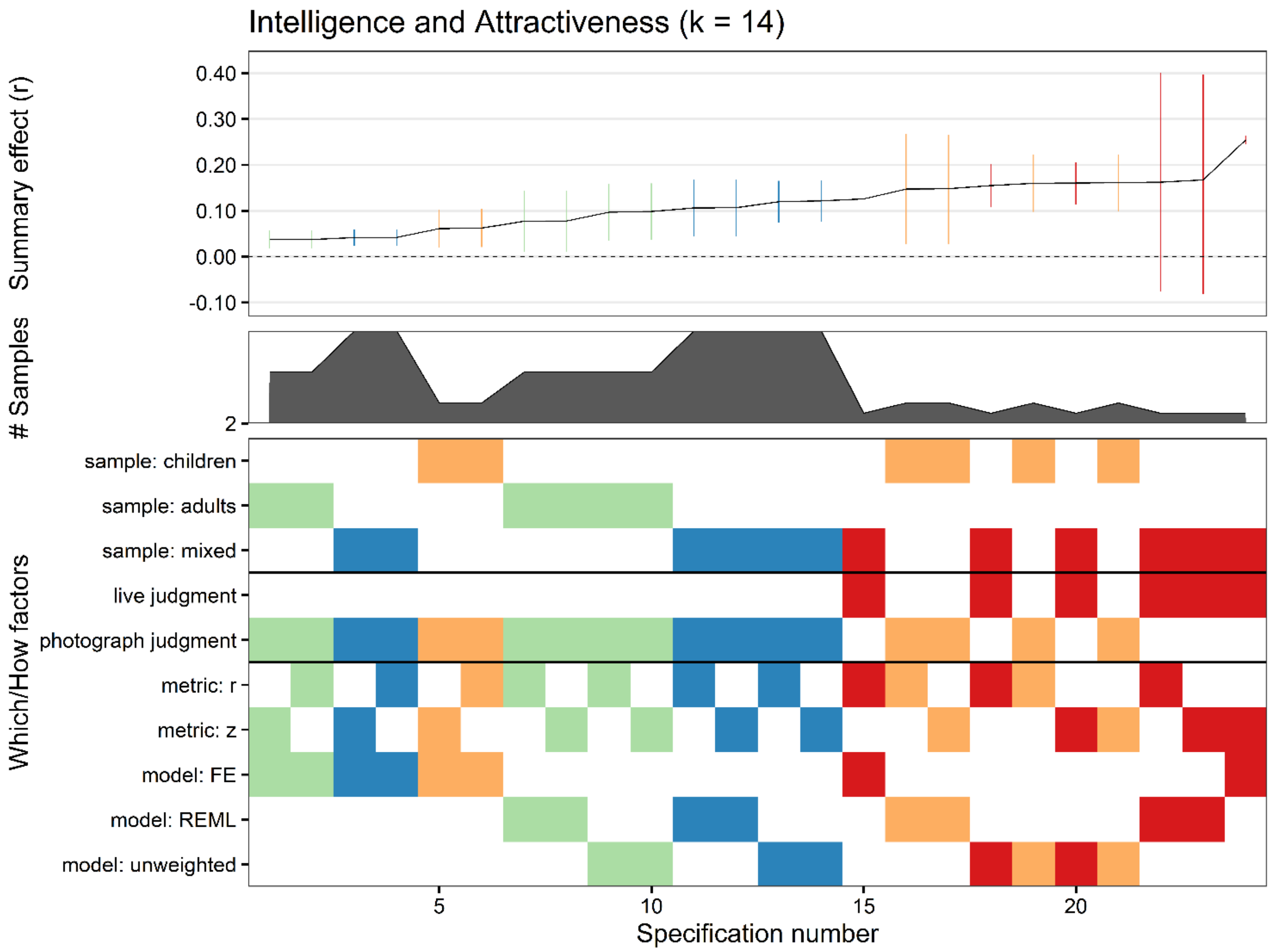
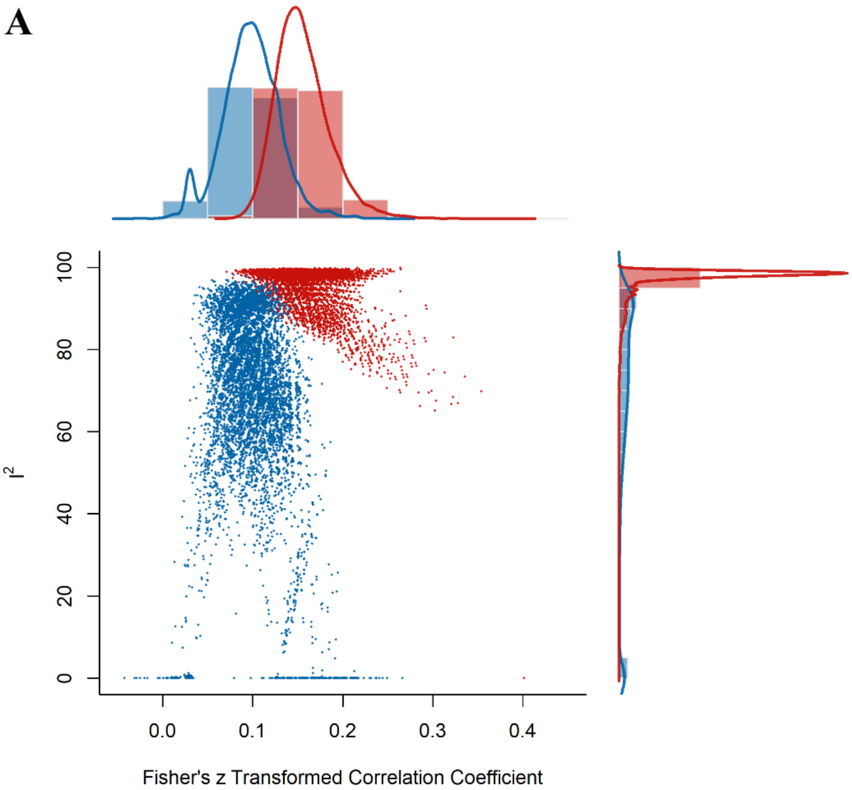
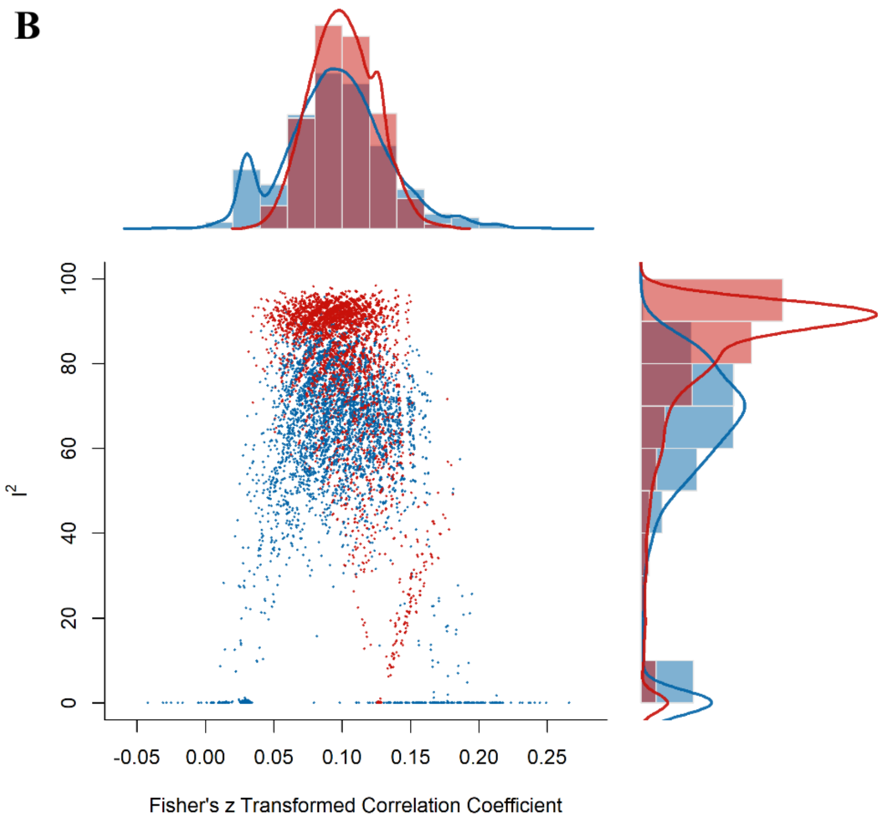


Figure 6. *GOSH-Plots. Each dot represents the summary effect of a random subset of studies. Panel A displays the full data (k = 14), which resulted in 16,383 combinations. Panel B shows the trimmed data (k = 13), yielding 8,191 different combinations. Subset estimations including at least one of the studies with larger effect sizes are visualized in red. Distributional densities are shown on the top (summary effects) and on the right side (I² values) of the plot.*

A



B



Discussion

The Dual Influence of Attractiveness: About the Health-Attractiveness Association, Social Reinforcement, and its Impact on Intelligence

The present meta-analysis provides evidence for a small yet robust association between intelligence as measured by psychometric IQ-tests and physical attractiveness. Despite the modest effect size, robustness analyses showed consistent results even after removing an influential outlier dataset. The correlation between intelligence and facial attractiveness aligns with previous theoretical perspectives and empirical findings. The "good genes" theory suggests that facial attractiveness serves as an indicator of genetic fitness and overall quality, including intelligence (Thornhill & Gangestad, 1993; Jones et al., 2001). There is a high degree of agreement both within and across cultures on what constitutes an attractive face (Cunningham et al., 1995; Langlois et al., 2000; Coetzee et al., 2014; Holzleitner et al., 2019), implying the presence of an underlying evolutionary mechanism for detecting "good genes". However, given the small correlation, it remains questionable whether physical attractiveness signals high intelligence or merely good health in general. The modest effect size suggests that the practical implications of this relationship may be limited. It is possible that attractiveness may serve as a more general indicator of good health, which can indirectly relate to cognitive functioning. Other factors beyond genetic fitness likely play a crucial role in determining intelligence, such as environmental influences (Nisbett et al., 2012; Pietschnig & Voracek, 2015), epigenetic factors (Haggarty et al., 2010; Yu et al., 2012; Kaminski et al., 2018), neurological and brain development factors (Deary et al., 2010; Menary et al., 2013; Schnack et al., 2015), motivation and personality traits (Chamorro-Premuzic & Furnham, 2003; O'Connor & Paunonen, 2007) cultural and societal factors (De Oliveira & Nisbett, 2017), and parenting and family environment (Turkheimer et al., 2003; Bates et al., 2013). These factors may interact with genetic predispositions in complex ways, shaping an individual's intellectual potential and cognitive development.

Both intelligence as well as physical attractiveness can be seen as indicators of underlying genetic fitness. For example, facial symmetry, a key component of attractiveness, is associated with good health and genetic quality (Thornhill & Gangestad 1993; Rhodes et al., 2001; Švegar, 2016). Facial symmetry is thought to reflect developmental stability, which is the ability for an organism to maintain stable development despite environmental stressors (Gangestad, 2022). Individuals with higher developmental stability are more likely to have symmetrical faces and are also more likely to have better overall health (Thornhill & Gangestad, 2006). Similarly, skin appearance has been identified as another important factor of facial attractiveness. Healthy skin is often associated with a balanced diet, adequate hydration, and a healthy lifestyle (Henderson et al., 2016; Pullar et al., 2017; Ryosuke et al., 2021). These lifestyle factors not only contribute to an attractive appearance but also promote overall health and well-being. The connection between attractiveness and health is particularly relevant when considering the relationship between intelligence and physical attractiveness. Individuals possessing higher cognitive abilities may show better health behaviors, which indirectly influences physical attractiveness. For instance, it was shown that a higher IQ is associated with overall lower rates of mortality (Betty et al., 2007). By contrast, lower IQ is linked with a wide range of negative health outcomes such as obesity (Goldberg et al., 2014), alcohol consumption (Sjölund et al., 2015), cigarette smoking (Weiser et al., 2010) and higher rates of coronary heart disease (Silventoinen et al., 2007; Yang et al., 2022). Thus, the interplay between health behaviors and cognitive abilities highlights how both can significantly impact physical attractiveness. Good health has also been shown to have a positive impact on cognitive development and performance (Bellisle, 2004; Haas, 2008). Children with better health tend to have higher IQ scores and better academic performance (Haas, 2008). Given the strong link between attractiveness and health, and the positive impact of good health on cognitive development and intellectual performance, it is plausible that the observed correlation between intelligence and physical attractiveness may be partially mediated by overall health. Attractive individuals, who are more likely to have good health, may also benefit from the cognitive advantages associated with a healthy lifestyle and optimal physical well-being.

Furthermore, the association between intelligence and physical attractiveness can be explained by the positive treatment of attractive individuals, which may influence the development and expression of cognitive abilities. This in turn could lead to a positive feedback loop that reinforces the correlation between IQ and physical attractiveness. Research suggests that individuals perceived as physically attractive often receive preferential treatment in various social and professional contexts, known as the "beauty premium" (Rosenblat, 2008; Deryugina & Shurchkov, 2015). This differential treatment can manifest in numerous ways, such as receiving more attention, praise, and opportunities for social interaction and advancement (Langlois et al., 2000). Such favorable treatment can have cascading effects on psychological well-being and self-esteem, providing a conducive environment for cognitive development and expression. For instance, individuals who experience positive reinforcement due to their attractiveness may feel more confident and socially competent, which can facilitate engagement in intellectually stimulating activities, such as academic pursuits or creative endeavors (Feingold, 1992; Lemay et al., 2010). Moreover, the heightened social visibility and influence enjoyed by attractive individuals may afford them greater opportunities for intellectual engagement and recognition, further enhancing their cognitive abilities. As a result, the positive treatment of attractive individuals not only perpetuates the attractiveness stereotype but also creates conditions conducive to the development and expression of cognitive abilities, thereby contributing to the observed correlation between IQ and physical attractiveness.

The Declining Correlation Between Attractiveness and Intelligence Over Time

The results revealed a declining trend in the strength of the attractiveness-intelligence correlation over time, with more recent studies reporting smaller effect sizes. The observed decline-effect is not unique to the domain of psychology and has been documented in various fields including medicine and biology (Schooler, 2011; de Bruin & Della Sala, 2015; Pietschnig et al., 2019). The decline in the attractiveness-intelligence association may be attributable to methodological improvements in more recent research, such as the use of larger and more representative samples, better control of confounding variables, or more robust measurement techniques. It is also plausible that earlier studies suffered from methodological limitations or publication biases that inflated the observed effects. As research practices improve and more rigorous standards are adopted, effect sizes may naturally decrease to reflect a more accurate estimation of the true underlying correlation (Schooler, 2011). In light of these considerations, more recently published studies may provide a more reliable picture of the association between intelligence and attractiveness.

Age-Dependent Differences in the Attractiveness-Intelligence Correlation: Stronger Association in Children Compared to Adults

The correlation between attractiveness and intelligence was stronger for children-only samples compared to adults-only samples. However, examination through moderator analysis revealed no significant differences between the age groups, indicating that the observed differences in correlation strength may be due to chance, rather than a true difference. Despite this, several factors might contribute to the observed trends. Children, particularly infants, naturally possess features aligned with the baby schema (Lorenz, 1943), such as large eyes and round faces, which evoke nurturing instincts and positive responses from adults (Glocker et al., 2009). However, the influence of the baby schema diminishes as children grow older. Attractiveness ratings also decline as face age increases, with the effect plateauing around 4.5

years of age (Luo et al., 2011). In the present meta-analysis, participants of the children samples were beyond the peak influence of the baby schema (age 7 and upwards). Intelligence is also more variable in childhood, which may contribute to the stronger observed correlation in this age group. This variability is supported by research showing that cognitive development in children undergoes rapid and diverse changes due to environmental, educational, and genetic factors (Blair, 2006; Tucker-Drob et al., 2011). As a result, the link between attractiveness and intelligence may be more pronounced in children because their cognitive abilities are still developing and being influenced by a wide range of external and internal factors. Moreover, compared to adults, children are less exposed to environmental toxins and lifestyle factors that could impair cognitive development and physical appearance. Factors such as smoking, alcohol consumption, and chronic stress, which can detrimentally affect both intelligence and physical attractiveness, are typically less prevalent in children's environments. This reduced exposure to environmental toxins may accentuate the observed correlation between intelligence and attractiveness in children.

Analysis of the adult samples showed a lower correlation between attractiveness and intelligence. The weaker correlation in adult samples could be attributed to the cumulative effect of environmental factors and lifestyle choices that may differentially impact intelligence and physical appearance. Age-related changes in physical appearance, such as wrinkles, changes in body composition, and hair loss can alter perceptions of attractiveness (Bertamini et al., 2013; Zebrowitz et al., 2013). These age-related changes may introduce variability that weakens the overall correlation between intelligence and attractiveness in adult samples. Over time, adults are also exposed to a wider range of environmental influences, such as education, occupation, lifestyle choices, and social interactions, which can differentially impact both cognitive abilities and physical attractiveness. As adults are more likely to encounter such environmental influences, the relationship between intelligence and attractiveness may become obscured. Additionally, as people age, the variance in life experiences such as education, occupation, and social interactions increases, further contributing to the divergence between cognitive abilities and physical appearance (Stern, 2002). Moreover, physical attractiveness is influenced by factors such as health and fitness levels, which can vary widely among adults due to differences in diet, exercise, and healthcare access. Adults with higher socioeconomic status may have better access to resources that enhance both cognitive development and physical appearance, such as higher education, healthier diets, and better healthcare (Gottfredson, 2004; Marmot, 2005). Conversely, lower socioeconomic status may result in limited access to these resources, potentially diminishing the correlation between intelligence and attractiveness. These complex and multifaceted influences likely dilute the direct correlation between intelligence and attractiveness observed in children. It is also conceivable that the correlation observed in adult samples reflects a more "adjusted" effect, as IQ tends to be more stable in adulthood and attractiveness ratings are no longer influenced by the baby schema or other childhood-specific factors. This stability in cognitive abilities and the absence of age-specific attractiveness biases may provide a clearer, albeit weaker, picture of the true relationship between intelligence and physical attractiveness.

Examining the Influence of Live and Photograph-Based Attractiveness Assessments on the Relationship Between Intelligence and Attractiveness

The analysis demonstrated a positive, although insignificant correlation between intelligence and attractiveness when attractiveness was assessed in live settings. It is notable that out of the 14 subsamples included in the analysis, only three used live assessments for attractiveness. This limited number of studies using live assessments may have contributed to the non-significant finding, as smaller sample sizes generally have lower statistical power to detect significant effects (Cohen, 2013; Button et al., 2013). In contrast, the correlation between intelligence and attractiveness was significantly positive, albeit slightly smaller, when attractiveness was assessed using photographs. Photographs provide a more standardized and controlled environment for assessing attractiveness, minimizing the influence of extraneous variables such as social interaction, body language, and other contextual factors present in live settings (Eagly et al., 1991; Feingold, 1992). This standardization allows for a clearer detection of the true effect of the attractiveness-intelligence relationship. Furthermore, photographs often capture static facial features, which have been shown to be more strongly associated with genetic quality and developmental stability compared to dynamic facial expressions (Rhodes, 2006; Thornhill & Gangestad, 1999). As a result, the association between intelligence and attractiveness may be more readily observable in photographs that emphasize stable facial characteristics. However, moderator analysis showed no difference between the types of attractiveness assessments. These findings indicate that other factors, such as sample characteristics or measurement techniques, might play a more crucial role in determining the strength of the observed correlation.

Publication Bias Analysis Corroborates the Validity of the Attractiveness-Intelligence Relationship

Overall, there were no strong indications for publication bias in the present meta-analysis. Most methods (i.e., funnel plot, Egger's regression test, the trim-and-fill method, the test of excess of significance, *p*-curve, and *p*-uniform*) suggested no evidence for publication bias. Only two methods (i.e., the rank correlation method and *p*-uniform) showed slight indications of publication bias, particularly when including the influential leverage study Kanazawa (2011) – Group A. The trim-and-fill analysis further highlighted the impact of this outlier, as its removal revealed potential missing studies on the left side of the funnel plot. Nonetheless as previously mentioned, most publication bias assessment methods show no substantial evidence for publication bias and *p*-hacking, suggesting that the observed association between intelligence and attractiveness is not influenced by selective reporting of results or questionable research practices. While some asymmetry was detected in the funnel plot of the trimmed analysis, the lack of significant excess of findings suggests that the observed results are driven by a genuine effect rather than publication bias. Furthermore, the observed asymmetry in the trimmed analysis highlights the impact of the leverage point (Kanazawa, 2011 – Group A). The analyses revealed that the inclusion of this outlier study may have masked some indications for publication bias in the full analysis. However, by removing this outlier, the true effect of the attractiveness-intelligence relationship becomes more precise. Even after removing the leverage study, the correlation remained positive, which underscores the stability of the attractiveness-intelligence association. The robustness of these findings is further supported by the consistent results across various publication bias detection methods and the significant positive correlation observed in the full dataset, even after adjusting for potential biases. In conclusion, the results of the present meta-analysis suggest that the relationship between intelligence and attractiveness stems from a true effect rather than an artifact of publication bias or methodological flaws.

The Stability of the Attractiveness-Intelligence Relationship: Findings from Specification Curve Analysis and Combinatorial Meta-Analyses

Overall, the specification curve analysis (SCA) and combinatorial meta-analyses (GOSH-plots) provided support for the robustness of the positive correlation between intelligence and attractiveness. All specifications of the SCA yielded a positive summary effect, indicating that the association between intelligence and attractiveness remained consistently positive. None of the specifications resulted in a null or negative effect, underscoring the robustness of the observed relationship. This consistency suggests that the association between intelligence and attractiveness is not highly sensitive to specific analytical choices, providing strong evidence that the effect is genuine and reliable. The robustness was also evident in the analysis of the trimmed dataset, which similarly revealed positive summary effects. This further supports the robustness of the positive association between intelligence and attractiveness, even when the outlier dataset was excluded. Overall, the results from the SCA reinforce the conclusion that more attractive people tend to be more intelligent, regardless of the specific analytical choices. This robustness across various specifications strengthens the validity of the findings and suggests that the observed association is not an artifact of particular analytic decisions.

Removing the outlier dataset resulted in a more homogenous set of studies, as indicated by the GOSH-plots, suggesting that the studies in the trimmed dataset have more consistent effect sizes, which leads to a more reliable and interpretable result. It also showed that in the trimmed dataset, the observed association between intelligence and attractiveness is not influenced by the extreme values of the leverage point. The consistency of the effect size across the remaining studies indicates a more stable and reliable association, reinforcing the validity of the findings. Without this leverage point, the effect appears more robust and consistent. This discrepancy underscores the atypical nature of the Kanazawa study, highlighting its disproportionate influence on the overall analysis. The remaining heterogeneity in the trimmed dataset may be due to inherent differences between the individual studies (i.e., IQ-tests, attractiveness ratings, sample characteristics, study design etc.). However, it is unclear which exact factors contribute to the remaining heterogeneity.

In summary, the analyses reinforce the conclusion that more attractive individuals tend to have higher intelligence, and this association remains robust across various analytical choices and after removing leverage dataset.

Limitations

Several limitations should be considered in the interpretation of these findings. Firstly, the influence of gender on the association between intelligence and attractiveness, as outlined in hypothesis 4, could not be examined due to insufficient data. Gender could impact societal perceptions and biases related to both intelligence as well as attractiveness. Some research also found a stronger correlation between general intelligence and attractiveness in men compared to women (Kanazawa, 2011). Forthcoming studies should aim for more detailed data collections to investigate the possible effect of gender-related differences. Furthermore, due to insufficient data, the potential moderating effect of using different IQ tests could not be examined. The limited number of studies reporting data on these factors precluded their inclusion in the meta-analysis.

One sample (Kanazawa, 2011 – Group A) showed an exceptionally high correlation between intelligence and attractiveness (r = .38) which had a large impact on overall results. Notably, attractiveness ratings within the sample by Kanazawa (2011) – Group A were provided by elementary school teachers, which potentially introduces a halo effect into the assessment process. The halo effect (Nisbett & Wilson, 1977), a cognitive bias where an overall positive impression of an individual influences the perception of other specific traits, may have played a role in the teachers' evaluations. Research over the past decades has indicated confirmation for a pronounced effect of the halo effect in various contexts and situations. For instance, Batres & Shiramizu (2023) found evidence that the attractiveness halo effect is apparent not only in Western cultures, but also among different cultures across the globe. Across 45 countries (in 11 world regions), attractiveness influenced ratings on personality characteristics, further supporting the underlying mechanisms of an attractiveness halo effect. Also, faces that are manipulated to look high in perceived intelligence are typically rated more attractive (Moore et al., 2011). Another factor that accounts for the exceptionally high correlation in this sample is the deployment of dichotomous attractiveness ratings, wherein a child was deemed attractive if rated as such at both ages 7 and 11. Dichotomous categorization forces raters to choose between extremes and leaves little room for nuanced distinctions. This may oversimplify the complex nature of attractiveness and inadvertently contribute to the inflation of correlation coefficients. Additionally, it is crucial to acknowledge that 62% of all participants in this sample were coded as attractive in the study by Kanazawa, 2011 – Group A. As such, the exceptional correlation observed in this particular sample should be interpreted with caution, considering the inherent limitations and potential biases associated with the method of assessment and the characteristics of the sample population.

Conclusion

The present meta-analysis provides evidence for a small robust correlation between intelligence and physical attractiveness. While the effect size is modest, the consistency of the findings across various analytical approaches lends credibility to the validity of this association even after excluding an influential outlier dataset. Beyond the direct associations with health and genetic fitness, physical attractiveness influences personality judgements and mating choices. In turn, this interplay may contribute to the observed correlation by affecting social interactions and mating opportunities. A person's personality contributes to the overall perception of an individual's desirability as a partner. Therefore, physical attractiveness may affect mating choices i) directly by highlighting good genes, but also ii) indirectly by signaling positive personality traits, such as intelligence. For example, symmetric faces of women tend to receive higher ratings for positive personality traits such as agreeableness, conscientiousness and extraversion, while asymmetric faces tend to be perceived as more neurotic, less agreeable and less conscientious (Fink et al., 2006). Attractive individuals also tend to exhibit higher levels of self-esteem and self-confidence and are therefore seen as more desirable partners (Bale & Archer, 2017; Chang, 2019). Humans possess a natural inclination to establish close relationships with individuals they perceive as physically attractive, a phenomenon termed "attractiveness-based affiliation motivation" (Lemay et al., 2010). This inclination is rooted in the perception that attractive individuals embody desirable interpersonal qualities such as kindness, warmth, and supportiveness (Lemay et al., 2010). The observed "what is beautiful is good" effect emerges as individuals project their own affiliation motivation onto their perceptions of the interpersonal attributes of attractive targets. As a result, humans tend to attribute greater receptiveness and responsiveness to attractive individuals, reflecting their own motivation to affiliate with them. Physical attractiveness may therefore not only reflect an individual's phenotypic quality, but also serves as an indicator for positive personality traits such as intelligence.

The attractiveness stereotype also extends to perceptions of intelligence. Research has consistently shown that individuals tend to ascribe higher intelligence to those who are physically attractive, a phenomenon known as the "attractiveness halo effect" (Eagly et al., 1991). Attractive individuals are often perceived as more competent and intelligent, regardless of their actual cognitive abilities (Zebrowitz et al., 2002; Kleisner et al., 2014). Additionally, the attractiveness halo is robust across different cultures and age groups, indicating a widespread bias where attractive individuals are assumed to possess more favorable traits, including higher intelligence (Langlois et al., 2000). This bias is so pervasive that it can influence educational and professional outcomes, with teachers and employers often favoring attractive individuals under the assumption of greater competence and intelligence (Clifford & Walster, 1973; Hosoda et al., 2003; Talamas et al., 2016). Furthermore, the correlation between intelligence and attractiveness is likely influenced by self-fulfilling prophecy effects, where perceived attractiveness shapes expectations and treatment of individuals, impacting their intellectual development and performance (Kleisner et al., 2014).

The precise extent to which offspring would inherit both intelligence and attractiveness remains questionable, given the nature of the heritance of traits. Genetic traits such as intelligence and physical attractiveness are considered polygenic traits, with numerous genes contributing to the overall expression of these traits (Davies et al., 2011; Hill et al., 2019; Genç et al., 2021; Fieder & Huber, 2023). The interaction of these genes can be complex, and having parents with high expressions of these traits does not guarantee that they will combine optimally or even be expressed in the offspring due to the recombination and variation occurring during genetic transmission. Previous research revealed that both fathers' and mothers' attractiveness significantly influence the facial attractiveness of daughters, but not the attractiveness of sons. No apparent link was found between mothers and sons in terms of facial traits, nor between fathers and daughters (Cornwell & Perrett, 2008). Consequently, a modest correlation between attractiveness and intelligence does not imply co-inheritance of both traits equally.

A potential genetic explanation for the correlation between IQ and physical attractiveness is the historical pattern of high-status men having greater reproductive success with multiple partners, often resulting in a higher number of offspring. This phenomenon, exemplified by the "Genghis Khan effect", where approximately 8% of men in the former Mongol Empire region carry a specific Y-chromosome lineage linked to Genghis Khan, underscores how high-status men could monopolize reproductive opportunities with numerous attractive partners (Zerjal et al., 2003; Derenko et al., 2007). Over generations, this selective mating pattern could lead to a genetic coupling of high intelligence and physical attractiveness within their descendants. However, it is challenging to make definitive statements about the genetic effects of this correlation due to the complex interplay of environmental, social, and cultural factors that also influence the expression of these traits. Historical polygyny might enhance the genetic link between IQ and attractiveness, however the actual prevalence and impact of such mating systems varied significantly across different societies and epochs, making it difficult to generalize these findings to the broader population. Also, the hypothesis that intelligence and attractiveness are correlated due to the intergenerational transmission of both traits through high-status individuals mating with attractive partners assumes that high-status historically equated to high intelligence. However, this assumption may not hold true in evolutionary history. High-status individuals in ancient societies often gained their positions through physical prowess, strategic alliances, or inheritance rather than through cognitive abilities alone (Betzig, 1993). For example, leaders and warriors might have achieved high status due to their physical strength, combat skills, or ability to command respect and loyalty, rather than their intellectual capabilities. Additionally, status could be influenced by factors such as kinship ties and social networks, which do not necessarily correlate with intelligence (Boone, 2017). Therefore, while intelligence and attractiveness are both heritable traits, the historical pathways through which high-status individuals achieved their positions suggest that high status did not automatically imply high intelligence. This weakens the argument that the correlation between intelligence and attractiveness is primarily driven by the mating patterns of high-status individuals. Since most men in our evolutionary history could only afford to support one wife, polygynous arrangements were typically reserved for the wealthy and powerful who had the resources to maintain multiple households. This linkage appears to persist in modern times, albeit to a lesser extent (De la Croix & Mariani, 2015; Ross et al., 2018).

In conclusion, the present meta-analysis provides robust evidence for a positive correlation between intelligence and physical attractiveness. However, it is essential to recognize that attractiveness is a multidimensional construct influenced by various factors beyond facial features, including health, overall well-being, and lifestyle choices. While facial attractiveness may serve as a cue for intelligence and good genes, it is not a reliable or strong predictor, and other factors likely play a more prominent role in determining cognitive abilities. Most importantly, the complexity of the correlation between IQ and attractiveness stems from the polygenic nature of these traits, leading to unpredictable outcomes in offspring despite paternal characteristics.

References

References marked with * are included in the meta-analysis.

Agthe, M., Spörrle, M., & Maner, J. K. (2010). Don't hate me because I'm beautiful: Anti-attractiveness bias in organizational evaluation and decision making. *Journal of Experimental Social Psychology*, *46*(6), 1151-1154. https://doi.org/10.1016/j.jesp.2010.05.007

Agthe, M., Strobel, M., Spörrle, M., Pfundmair, M., & Maner, J. K. (2016). On the borders of harmful and helpful beauty biases: The biasing effects of physical attractiveness depend on sex and ethnicity. *Evolutionary Psychology*, *14*(2), 1474704916653968. https://doi.org/10.1177/1474704916653968

Archer, J. (2019). The reality and evolutionary significance of human psychological sex differences. *Biological Reviews*, *94*(4), 1381-1415. https://doi.org/10.1111/brv.12507

Arden, R., Gottfredson, L. S., & Miller, G. (2009). Does a fitness factor contribute to the association between intelligence and health outcomes? Evidence from medical abnormality counts among 3654 US Veterans. *Intelligence*, *37*(6), 581-591. https://doi.org/10.1016/j.intell.2009.03.008

Arnocky, S., Sunderani, S., Miller, J. L., & Vaillancourt, T. (2012). Jealousy mediates the relationship between attractiveness comparison and females' indirect aggression. *Personal Relationships*, *19*(2), 290-303. https://doi.org/10.1111/j.1475-6811.2011.01362.x

Bale, C., & Archer, J. (2013). Self-perceived attractiveness, romantic desirability and self-esteem: A mating sociometer perspective. *Evolutionary Psychology*, *11*(1), 147470491301100107. https://doi.org/10.1177/147470491301100107

Bartels, M., Rietveld, M. J., Van Baal, G. C. M., & Boomsma, D. I. (2002). Genetic and environmental influences on the development of intelligence. *Behavior genetics*, *32*, 237-249. https://doi.org/10.1023/A:1019772628912

Bates, T. C., Lewis, G. J., & Weiss, A. (2013). Childhood socioeconomic status amplifies genetic effects on adult intelligence. *Psychological science*, *24*(10), 2111-2116. https://doi.org/10.1177/0956797613488394

Batres, C., & Shiramizu, V. (2023). Examining the “attractiveness halo effect” across cultures. *Current Psychology*, *42*(29), 25515-25519. https://doi.org/10.1007/s12144-022-03575-0

Baumeister, R. F., Catanese, K. R., & Vohs, K. D. (2001). Is there a gender difference in strength of sex drive? Theoretical views, conceptual distinctions, and a review of relevant evidence. *Personality and social psychology review*, *5*(3), 242-273. https://doi.org/10.1207/S15327957PSPR0503_5

Bellisle, F. (2004). Effects of diet on behaviour and cognition in children. *British Journal of nutrition*, *92*(S2), S227-S232. https://doi.org/10.1079/BJN20041171

Bergman, L. R., Corovic, J., Ferrer-Wreder, L., & Modig, K. (2014). High IQ in early adolescence and career success in adulthood: Findings from a Swedish longitudinal study. *Research in Human Development*, *11*(3), 165-185. https://doi.org/10.1080/15427609.2014.936261

Bertamini, M., Rampone, G., Makin, A. D., & Jessop, A. (2019). Symmetry preference in shapes, faces, flowers and landscapes. *PeerJ*, *7*, e7078. https://doi.org/10.7717/peerj.7078

Betzig, L. (1993). Sex, succession, and stratification in the first six civilizations: How powerful men reproduced, passed power on to their sons, and used power to defend their wealth, women, and children. In L. Ellis (Ed.), *Social stratification and socioeconomic inequality: Vol. 1. A comparative biosocial analysis* (pp. 37–74).

Betzig, L. (2012). Means, variances, and ranges in reproductive success: comparative evidence. *Evolution and Human Behavior*, *33*(4), 309-317. https://doi.org/10.1016/j.evolhumbehav.2011.10.008

Betzig, L. L. (1986). *Despotism and differential reproduction: A Darwinian view of history*. Aldine Publishing Co.

Blair, C. (2006). How similar are fluid cognition and general intelligence? A developmental neuroscience perspective on fluid cognition as an aspect of human cognitive ability. *Behavioral and Brain Sciences*, *29*(2), 109-125. https://doi.org/10.1017/S0140525X06009034

Boone, J. L. (2017). Competition, conflict, and the development of social hierarchies. In *Evolutionary ecology and human behavior* (pp. 301-338). Routledge.

Boothroyd, L. G., Scott, I., Gray, A. W., Coombes, C. I., & Pound, N. (2013). Male facial masculinity as a cue to health outcomes. *Evolutionary psychology*, *11*(5), 147470491301100508. https://doi.org/10.1177/1474704913011005

Border, R., Athanasiadis, G., Buil, A., Schork, A. J., Cai, N., Young, A. I., ... & Zaitlen, N. A. (2022). Cross-trait assortative mating is widespread and inflates genetic correlation estimates. *Science*, *378*(6621), 754-761. https://doi.org/10.1126/science.abo2059

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2021). Introduction to meta-analysis. *John Wiley & Sons*.

Bornstein, M. H., Ferdinandsen, K., & Gross, C. G. (1981). Perception of symmetry in infancy. *Developmental psychology*, *17*(1), 82. https://doi.org/10.1037/0012-1649.17.1.82

Bouchard Jr, T. J., Lykken, D. T., McGue, M., Segal, N. L., & Tellegen, A. (1990). Sources of human psychological differences: The Minnesota study of twins reared apart. Science, 250(4978), 223-228. https://doi.org/10.1126/science.2218526

Bulczak, G., & Gugushvili, A. (2023). Physical attractiveness and cardiometabolic risk. *American Journal of Human Biology*, *35*(8), e23895. https://doi.org/10.1002/ajhb.23895

Buss, D. M. (1989). Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. *Behavioral and brain sciences*, *12*(1), 1-14. https://doi.org/10.1017/S0140525X00023992

Buss, D. M. (1994). The evolution of desire: Strategies of human mating. New York: BasicBooks.

Buss, D. M., & Schmitt, D. P. (2019). Mate preferences and their behavioral manifestations. *Annual review of psychology*, *70*, 77-110. https://doi.org/10.1146/annurev-psych-010418-103408

Buss, D. M., & Shackelford, T. K. (2008). Attractive women want it all: Good genes, economic investment, parenting proclivities, and emotional commitment. *Evolutionary Psychology*, *6*(1). https://doi.org/10.1177/147470490800600116

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature reviews neuroscience*, *14*(5), 365-376. https://doi.org/10.1038/nrn3475

Chamorro-Premuzic, T., & Furnham, A. (2003). Personality predicts academic performance: Evidence from two longitudinal university samples. *Journal of research in personality*, *37*(4), 319-338. https://doi.org/10.1016/S0092-6566(02)00578-0

Chance, M., Jackson Ph D, A. T., & Frame, M. (2023). Pretty privilege at work: the influence of physical attractiveness on hiring and rating decisions.

Chang, W. L. (2019). Does beauty matter? Exploring the relationship between self-consciousness and physical attractiveness. *Kybernetes*, *48*(3), 362-384. https://doi.org/10.1108/K-12-2017-0494

Choi, K. Y. (2015). Analysis of facial asymmetry. Archives of craniofacial surgery, 16(1), 1. https://doi.org/10.7181/acfs.2015.16.1.1

Chu, S., Hardaker, R., & Lycett, J. E. (2007). Too good to be 'true'? The handicap of high socio-economic status in attractive males. *Personality and Individual Differences, 42*(7), 1291-1300. https://doi.org/10.1016/j.paid.2006.10.007

Čivre, Ž., Knežević, M., Baruca, P. Z., & Fabjan, D. (2013). Facial attractiveness and stereotypes of hotel guests: An experimental research. *Tourism management*, *36*, 57-65. https://doi.org/10.1016/j.tourman.2012.11.004

Clifford, M. M., & Walster, E. (1973). The effect of physical attractiveness on teacher expectations. *Sociology of education*, 248-258. https://doi.org/10.2307/2112099

Coetzee, V., Greeff, J. M., Stephen, I. D., & Perrett, D. I. (2014). Cross-cultural agreement in facial attractiveness preferences: The role of ethnicity and gender. *PloS one*, *9*(7), e99629. https://doi.org/10.1371/journal.pone.0099629

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. routledge.

Cole, J. B., Manyama, M., Larson, J. R., Liberton, D. K., Ferrara, T. M., Riccardi, S. L., ... & Spritz, R. A. (2017). Human facial shape and size heritability and genetic correlations. *Genetics*, *205*(2), 967-978. *https://doi.org/10.1534/genetics.116.193185*

*Colladon, A. F., Grippa, F., Battistoni, E., Gloor, P. A., & Bella, A. L. (2018). What makes you popular: Beauty, personality or intelligence? International Journal of Entrepreneurship and Small Business, 35(2), 162–186. https://doi.org/10.1504/IJESB.2018.094967

Cooper, H., Hedges, L. V., & Valentine, J. C. (2019). The handbook of research synthesis and meta-analysis. *Russell Sage Foundation*.

Cornwell, R. E., & Perrett, D. I. (2008). Sexy sons and sexy daughters: the influence of parents' facial characteristics on offspring. *Animal Behaviour*, *76*(6), 1843-1853. https://doi.org/10.1016/j.anbehav.2008.07.031

Čukić, I., Brett, C. E., Calvin, C. M., Batty, G. D., & Deary, I. J. (2017). Childhood IQ and survival to 79: Follow-up of 94% of the Scottish Mental Survey 1947. *Intelligence*, *63*, 45-50. https://doi.org/10.1016/j.intell.2017.05.002

Cunningham, M. R., Roberts, A. R., Barbee, A. P., Druen, P. B., & Wu, C. H. (1995). " Their ideas of beauty are, on the whole, the same as ours": Consistency and variability in the cross-cultural perception of female physical attractiveness. *Journal of personality and social psychology*, *68*(2), 261. https://doi.org/10.1037/0022-3514.68.2.261

Davies, G., Tenesa, A., Payton, A., Yang, J., Harris, S. E., Liewald, D., ... & Deary, I. J. (2011). Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Molecular psychiatry*, *16*(10), 996-1005. https://doi.org/10.1038/mp.2011.85

de Bruin, A., & Della Sala, S. (2015). The decline effect: How initially strong results tend to decrease over time. *Cortex*, *73*, 375-377. https://doi.org/10.1016/j.cortex.2015.05.025

De la Croix, D., & Mariani, F. (2015). From polygyny to serial monogamy: a unified theory of marriage institutions. *The Review of Economic Studies*, *82*(2), 565-607. https://doi.org/10.1093/restud/rdv001

De Oliveira, S., & Nisbett, R. E. (2017). Culture changes how we think about thinking: From "Human Inference" to "Geography of Thought". *Perspectives on Psychological Science*, *12*(5), 782-790. https://doi.org/10.1177/1745691617702718

Deary, I. J., Penke, L., & Johnson, W. (2010). The neuroscience of human intelligence differences. *Nature reviews neuroscience*, *11*(3), 201-211. https://doi.org/10.1038/nrn2793

Deming, D. J. (2017). The growing importance of social skills in the labor market. *The quarterly journal of economics*, *132*(4), 1593-1640. https://doi.org/10.1093/qje/qjx022

Derenko, M. V., Malyarchuk, B. A., Wozniak, M., Denisova, G. A., Dambueva, I. K., Dorzhu, C. M., ... & Zakharov, I. A. (2007). Distribution of the male lineages of Genghis Khan's descendants in northern Eurasian populations. *Russian Journal of Genetics*, *43*, 334-337. https://doi.org/10.1134/S1022795407030179

Deryugina, T., & Shurchkov, O. (2015). Now you see it, now you don't: The vanishing beauty premium. *Journal of Economic Behavior & Organization*, *116*, 331-345. https://doi.org/10.1016/j.jebo.2015.05.007

Devlin, B., Daniels, M., & Roeder, K. (1997). The heritability of IQ. *Nature*, *388*(6641), 468-471. https://doi.org/10.1038/41319

Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of personality and social psychology*, *24*(3), 285. https://doi.org/10.1037/h0033731

Duval, S., & Tweedie, R. (2000). Trim and fill: a simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455-463. https://doi.org/10.1111/j.0006-341X.2000.00455.x

Eastwick, P. W., & Finkel, E. J. (2008). Sex differences in mate preferences revisited: Do people know what they initially desire in a romantic partner?. *Journal of personality and social psychology*, *94*(2), 245. https://doi.org/10.1037/0022-3514.94.2.245

Eaves, L. J., Last, K. A., Young, P. A., & Martin, N. G. (1978). Model-fitting approaches to the analysis of human behaviour. *Heredity*, *41*(3), 249-320. https://doi.org/10.1038/hdy.1978.101

Feingold, A. (1988). Matching for attractiveness in romantic partners and same-sex friends: A meta-analysis and theoretical critique. *Psychological bulletin*, *104*(2), 226. https://doi.org/10.1037/0033-2909.104.2.226

Feingold, A. (1992). Gender differences in mate selection preferences: a test of the parental investment model. *Psychological bulletin*, *112*(1), 125. https://doi.org/10.1037/0033-2909.112.1.125

Fieder, M., & Huber, S. (2023). Facial attractiveness is only weakly linked to genome–wide heterozygosity. *Frontiers in Psychology*, *14*, 1009962. https://doi.org/10.3389/fpsyg.2023.1009962

Fisher, H. E. (1989). Evolution of human serial pairbonding. *American Journal of Physical Anthropology*, *78*(3), 331-354. https://doi.org/10.1002/ajpa.1330780303

Fuxjager, M. J., Foufopoulos, J., Diaz-Uriarte, R., & Marler, C. A. (2011). Functionally opposing effects of testosterone on two different types of parasite: implications for the immunocompetence handicap hypothesis. *Functional Ecology*, *25*(1), 132-138. https://doi.org/10.1111/j.1365-2435.2010.01784.x

Gangestad, S. W. (2022). Developmental instability, fluctuating asymmetry, and human psychological science. *Emerging Topics in Life Sciences*, *6*(3), 311-322. https://doi.org/10.1042/ETLS20220025

Gangestad, S. W., & Scheyd, G. J. (2005). The evolution of human physical attractiveness. *Annu. Rev. Anthropol.*, *34*, 523-548. https://doi.org/10.1146/annurev.anthro.33.070203.143733

Gangestad, S. W., & Scheyd, G. J. (2005). The evolution of human physical attractiveness. *Annu. Rev. Anthropol.*, *34*(1), 523-548. https://doi.org/10.1146/annurev.anthro.33.070203.143733

Gangestad, S. W., & Simpson, J. A. (2000). The evolution of human mating: Trade-offs and strategic pluralism. *Behavioral and brain sciences*, *23*(4), 573-587. https://doi.org/10.1017/S0140525X0000337X

Gangestad, S. W., & Thornhill, R. (2003). Facial masculinity and fluctuating asymmetry. *Evolution and Human Behavior*, *24*(4), 231-241. https://doi.org/10.1016/S1090-5138(03)00017-5

Gangestad, S. W., & Thornhill, R. (2003). Facial masculinity and fluctuating asymmetry. *Evolution and Human Behavior*, *24*(4), 231-241. https://doi.org/10.1016/S1090-5138(03)00017-5

*Gangestad, S. W., Thornhill, R., & Garver-Apgar, C. E. (2010). Men's facial masculinity predicts changes in their female partners' sexual interests across the ovulatory cycle, whereas men's intelligence does not. Evolution and Human Behavior, 31(6), 412–424. https://doi.org/10.1016/j.evolhumbehav.2010.06.001

Gangestad, S. W., Thornhill, R., & Yeo, R. A. (1994). Facial attractiveness, developmental stability, and fluctuating asymmetry. *Ethology and Sociobiology*, *15*(2), 73-85. https://doi.org/10.1016/0162-3095(94)90018-3

Gavrilets, S. (2012). Human origins and the transition from promiscuity to pair-bonding. *Proceedings of the National Academy of Sciences*, *109*(25), 9923-9928. https://doi.org/10.1073/pnas.1200717109

Genç, E., Schlüter, C., Fraenz, C., Arning, L., Metzen, D., Nguyen, H. P., ... & Ocklenburg, S. (2021). Polygenic scores for cognitive abilities and their association with different aspects of general intelligence—a deep phenotyping approach. *Molecular Neurobiology*, *58*(8), 4145-4156.

Gignac, G. E., Darbyshire, J., & Ooi, M. (2018). Some people are attracted sexually to intelligence: A psychometric evaluation of sapiosexuality. *Intelligence*, *66*, 98-111. https://doi.org/10.1016/j.intell.2017.11.009

Gildersleeve, K., Haselton, M. G., & Fales, M. R. (2014). Do women's mate preferences change across the ovulatory cycle? A meta-analytic review. *Psychological bulletin*, *140*(5), 1205. https://doi.org/10.1037/a0035438

Glocker, M. L., Langleben, D. D., Ruparel, K., Loughead, J. W., Gur, R. C., & Sachser, N. (2009). Baby schema in infant faces induces cuteness perception and motivation for caretaking in adults. *Ethology*, *115*(3), 257-263. https://doi.org/10.1111/j.1439-0310.2008.01603.x

González-Saldivar, G., Rodríguez-Gutiérrez, R., Ocampo-Candiani, J., González-González, J. G., & Gómez-Flores, M. (2017). Skin manifestations of insulin resistance: from a biochemical stance to a clinical diagnosis and management. *Dermatology and Therapy*, *7*, 37-51. https://doi.org/10.1007/s13555-016-0160-3

Gottfredson, L. S., & Deary, I. J. (2004). Intelligence predicts health and longevity, but why?. *Current Directions in Psychological Science*, *13*(1), 1-4. https://doi.org/10.1111/j.0963-7214.2004.01301001.x

Graham, J. H., & Özener, B. (2016). Fluctuating asymmetry of human populations: A review. *Symmetry*, *8*(12), 154. https://doi.org/10.3390/sym8120154

Grammer, K., & Thornhill, R. (1994). Human (Homo sapiens) facial attractiveness and sexual selection: the role of symmetry and averageness. *Journal of comparative psychology*, *108*(3), 233. https://doi.org/10.1037/0735-7036.108.3.233

Guadagno, R. E., & Sagarin, B. J. (2010). Sex differences in jealousy: An evolutionary perspective on online infidelity. *Journal of Applied Social Psychology*, *40*(10), 2636-2655. https://doi.org/10.1111/j.1559-1816.2010.00674.x

Gubbels Bupp, M. R., & Jorgensen, T. N. (2018). Androgen-induced immunosuppression. *Frontiers in immunology*, *9*, 794. https://doi.org/10.3389/fimmu.2018.00794

Ha, T., Van Den Berg, J. E., Engels, R. C., & Lichtwarck-Aschoff, A. (2012). Effects of attractiveness and status in dating desire in homosexual and heterosexual men and women. *Archives of Sexual Behavior*, *41*, 673-682. https://doi.org/10.1007/s10508-011-9855-9

Haas, S. (2008). Trajectories of functional health: The 'long arm'of childhood health and socioeconomic factors. *Social science & medicine*, *66*(4), 849-861. https://doi.org /10.1016/j.socscimed.2007.11.004

Haggarty, P., Hoad, G., Harris, S. E., Starr, J. M., Fox, H. C., Deary, I. J., & Whalley, L. J. (2010). Human intelligence and polymorphisms in the DNA methyltransferase genes involved in epigenetic marking. *PloS one*, *5*(6), e11329. https://doi.org/10.1371/journal.pone.0011329

*Härtel, T. M., Leckelt, M., Grosz, M. P., Küfner, A. C. P., Geukes, K., & Back, M. D. (2023). Pathways From Narcissism to Leadership Emergence in Social Groups. European Journal of Personality, 37(1), 72–94. Scopus. https://doi.org/10.1177/08902070211046266

Haworth, C. M., Wright, M. J., Luciano, M., Martin, N. G., de Geus, E. J., van Beijsterveldt, C. E., ... & Plomin, R. (2010). The heritability of general cognitive ability increases linearly from childhood to young adulthood. *Molecular psychiatry*, *15*(11), 1112-1120. https://doi.org/10.1038/mp.2009.55

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor economics*, *24*(3), 411-482. https://doi.org/10.1086/504455

Heilman, M. E., & Stopeck, M. H. (1985). Attractiveness and corporate success: Different causal attributions for males and females. *Journal of Applied Psychology*, *70*(2), 379. https:///doi.org/10.1037/0021-9010.70.2.379

Henderson, A. J., Holzleitner, I. J., Talamas, S. N., & Perrett, D. I. (2016). Perception of health from facial cues. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1693), 20150380. https://doi.org/10.1098/rstb.2015.0380

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *Bmj*, *327*(7414), 557-560. https://doi.org/10.1136/bmj.327.7414.557

Hilger, K., Spinath, F. M., Troche, S., & Schubert, A. L. (2022). The biological basis of intelligence: Benchmark findings. *Intelligence*, *93*, 101665. https://doi.org/10.1016/j.intell.2022.101665

Hill, W. D., Marioni, R. E., Maghzian, O., Ritchie, S. J., Hagenaars, S. P., McIntosh, A. M., ... & Deary, I. J. (2019). A combined analysis of genetically correlated traits identifies 187 loci and a role for neurogenesis and myelination in intelligence. *Molecular psychiatry*, *24*(2), 169-181. https://doi.org/10.1038/s41380-017-0001-5

Holzleitner, I. J., Lee, A. J., Hahn, A. C., Kandrik, M., Bovet, J., Renoult, J. P., ... & Jones, B. C. (2019). Comparing theory-driven and data-driven attractiveness models using images of real women's faces. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(12), 1589. https://doi.org/10.1037/xhp0000685

Hosoda, M., Stone-Romero, E. F., & Coats, G. (2003). The effects of physical attractiveness on job-related outcomes: A meta-analysis of experimental studies. *Personnel psychology*, *56*(2), 431-462. https://doi.org/10.1111/j.1744-6570.2003.tb00157.x

Hu, B., Shen, N., Li, J. J., Kang, H., Hong, J., Fletcher, J., ... & Lu, Q. (2019). Genome-wide association study reveals sex-specific genetic architecture of facial attractiveness. *PLoS Genetics*, *15*(4), e1007973. https://doi.org/10.1371/journal.pgen.1007973

Huang, Y., Xue, X., Spelke, E., Huang, L., Zheng, W., & Peng, K. (2018). The aesthetic preference for symmetry dissociates from early-emerging attention to symmetry. *Scientific reports*, *8*(1), 6263. https://doi.org/10.1038/s41598-018-24558-x

Hunt, E. (2010). Human intelligence. Cambridge University Press.

Johnson, S. K., Podratz, K. E., Dipboye, R. L., & Gibbons, E. (2010). Physical attractiveness biases in ratings of employment suitability: Tracking down the "beauty is beastly" effect. *The Journal of social psychology*, *150*(3), 301-318. https://doi.org/10.1080/00224540903365414

Jones, B. C., DeBruine, L. M., & Little, A. C. (2007). The role of symmetry in attraction to average faces. Perception & psychophysics, 69, 1273-1277. https://doi.org/10.3758/BF03192944

Jones, B. C., Little, A. C., Penton-Voak, I. S., Tiddeman, B. P., Burt, D. M., & Perrett, D. I. (2001). Facial symmetry and judgements of apparent health: Support for a "good genes" explanation of the attractiveness–symmetry relationship. *Evolution and human behavior*, *22*(6), 417-429. https://doi.org/10.1016/S1090-5138(01)00083-6

Jones, D., & Hill, K. (1993). Criteria of facial attractiveness in five populations. *Human Nature*, *4*(3), 271-296. https://doi.org/10.1007/BF02692202

Judge, T. A., Colbert, A. E., & Ilies, R. (2004). Intelligence and leadership: a quantitative review and test of theoretical propositions. *Journal of applied psychology*, *89*(3), 542. https://psycnet.apa.org/doi/10.1037/0021-9010.89.3.542

*Judge, T. A., Hurst, C., & Simon, L. S. (2009). Does It Pay to Be Smart, Attractive, or Confident (or All Three)? Relationships Among General Mental Ability, Physical Attractiveness, Core Self-Evaluations, and Income. Journal of Applied Psychology, 94(3), 742–755. Scopus. https://doi.org/10.1037/a0015497

Kaminski, J. A., Schlagenhauf, F., Rapp, M., Awasthi, S., Ruggeri, B., Deserno, L., ... & IMAGEN consortium. (2018). Epigenetic variance in dopamine D2 receptor: a marker of IQ malleability?. *Translational Psychiatry*, *8*(1), 169. https://doi.org/10.1038/s41398-018-0222-7

Kanazawa, S., & Kovar, J. L. (2004). Why beautiful people are more intelligent. *Intelligence*, *32*(3), 227-243. https://doi.org/10.1016/j.intell.2004.03.003

*Kanazawa, S. (2011). Intelligence and physical attractiveness. Intelligence, 39(1), 7–14. https://doi.org/10.1016/j.intell.2010.11.003

Kanazawa, S., & Still, M. C. (2018). Is there really a beauty premium or an ugliness penalty on earnings?. *Journal of Business and Psychology*, *33*(2), 249-262. https://doi.org/10.1007/s10869-017-9489-6

*Kleisner, K., Chvátalová, V., & Flegr, J. (2014). Perceived Intelligence Is Associated with Measured Intelligence in Men but Not Women. PLOS ONE, 9(3), e81237. https://doi.org/10.1371/journal.pone.0081237

Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average. *Psychological science*, *1*(2), 115-121. https://doi.org/10.1111/j.1467-9280.1990.tb00079

*Lee, A. J., Hibbs, C., Wright, M. J., Martin, N. G., Keller, M. C., & Zietsch, B. P. (2017). Assessing the accuracy of perceptions of intelligence based on heritable facial features. Intelligence, 64, 1–8. https://doi.org/10.1016/j.intell.2017.06.002

Lemay Jr, E. P., Clark, M. S., & Greenberg, A. (2010). What is beautiful is good because what is beautiful is desired: Physical attractiveness stereotyping as projection of interpersonal goals. *Personality and Social Psychology Bulletin*, *36*(3), 339-353. https://doi.org/10.1177/0146167209359700

Lephart, E. D. (2018). A review of the role of estrogen in dermal aging and facial attractiveness in women. *Journal of cosmetic dermatology*, *17*(3), 282-288. https://doi.org/10.1111/jocd.12508

Li, N. P., Bailey, J. M., Kenrick, D. T., & Linsenmeier, J. A. (2002). The necessities and luxuries of mate preferences: testing the tradeoffs. *Journal of personality and social psychology*, *82*(6), 947. https://doi.org/10.1037/0022-3514.82.6.947

Li, N. P., Valentine, K. A., & Patel, L. (2011). Mate preferences in the US and Singapore: A cross-cultural test of the mate preference priority model. *Personality and Individual Differences*, *50*(2), 291-294. https://doi.org/10.1016/j.paid.2010.10.005

Li, Y., Zhang, C., & Laroche, M. (2019). Is beauty a premium? A study of the physical attractiveness effect in service encounters. *Journal of Retailing and Consumer Services*, *50*, 215-225. https://doi.org/10.1016/j.jretconser.2019.04.016

Lorenz, K. (1943). Die angeborenen formen möglicher erfahrung. *Zeitschrift für Tierpsychologie*, *5*(2), 235-409. https://doi.org/10.1111/j.1439-0310.1943.tb00655.x

Lorenzo, G. L., Biesanz, J. C., & Human, L. J. (2010). What is beautiful is good and more accurately understood: Physical attractiveness and accuracy in first impressions of personality. *Psychological science*, *21*(12), 1777-1782. https://doi.org/10.1177/0956797610388048

Lovejoy, C. O. (1981). The origin of man. *Science*, *211*(4480), 341-350. https://doi.org/10.1126/science.211.4480.341

Luo, L. Z., Li, H., & Lee, K. (2011). Are children's faces really more appealing than those of adults? Testing the baby schema hypothesis beyond infancy. *Journal of experimental child psychology*, *110*(1), 115-124. https://doi.org/10.1016/j.jecp.2011.04.002

Luo, S., & Zhang, G. (2009). What leads to romantic attraction: Similarity, reciprocity, security, or beauty? Evidence from a speed-dating study. *Journal of personality*, *77*(4), 933-964. https://doi.org/10.1111/j.1467-6494.2009.00570.x

Marcus, D. K., & Miller, R. S. (2003). Sex differences in judgments of physical attractiveness: A social relations analysis. *Personality and Social Psychology Bulletin*, *29*(3), 325-335. https://doi.org/10.1177/0146167202250193

Marlowe, F. (2000). Paternal investment and the human mating system. *Behavioural processes*, *51*(1-3), 45-61. https://doi.org/10.1016/S0376-6357(00)00118-2

Marlowe, F. W. (2003). The mating system of foragers in the standard cross-cultural sample. *Cross-Cultural Research*, *37*(3), 282-306. https://doi.org/10.1177/1069397103254008

Marmot, M. (2005). Social determinants of health inequalities. *The lancet*, *365*(9464), 1099-1104. https://doi.org/10.1016/S0140-6736(05)71146-6

McClintock, E. A. (2014). Beauty and status: The illusion of exchange in partner selection?. *American sociological review*, *79*(4), 575-604. https://doi.org/10.1177/0003122414536391

McGovern, R. J., Neale, M. C., & Kendler, K. S. (1996). The independence of physical attractiveness and symptoms of depression in a female twin population. *The Journal of psychology*, *130*(2), 209-219. https://doi.org/10.1080/00223980.1996.9915002

McShane, B. B., Böckenholt, U., & Hansen, K. T. (2016). Adjusting for publication bias in meta-analysis: An evaluation of selection methods and some cautionary notes. *Perspectives on Psychological Science*, *11*(5), 730-749. https://doi.org/10.1177/1745691616662243

Møller, A. P. (1998). Developmental instability as a general measure of stress. In *Advances in the Study of Behavior* (Vol. 27, pp. 181-213). Academic Press. https://doi.org/10.1016/S0065-3454(08)60365-4

Møller, A. P., & Thornhill, R. (1998). Bilateral symmetry and sexual selection: a meta-analysis. *The American Naturalist*, *151*(2), 174-192. https://doi.org/10.1086/286110

Moore, F. R., Filippou, D., & Perrett, D. I. (2011). Intelligence and attractiveness in the face: Beyond the attractiveness halo effect. *Journal of Evolutionary Psychology*, *9*(3), 205-217. https://doi.org/10.1556/jep.9.2011.3.2

Murdock, G. P. (1981). *Atlas of world cultures*. University of Pittsburgh Pre.

Murtza, M. H., Gill, S. A., Aslam, H. D., & Noor, A. (2021). Intelligence quotient, job satisfaction, and job performance: The moderating role of personality type. *Journal of Public Affairs*, *21*(3), e2318. https://doi.org/10.1002/pa.2318

Nedelec, J. L., & Beaver, K. M. (2014). Physical attractiveness as a phenotypic marker of health: an assessment using a nationally representative sample of American adults.

Evolution and Human Behavior, *35*(6), 456-463. https://doi.org/10.1016/j.evolhumbehav.2014.06.004

Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of personality and social psychology*, *35*(4), 250. https://doi.org/10.1037/0022-3514.35.4.250

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: new findings and theoretical developments. *American psychologist*, *67*(2), 130. https://doi.org/10.1037/a0026699

Nowak, J., Pawłowski, B., Borkowska, B., Augustyniak, D., & Drulis-Kawa, Z. (2018). No evidence for the immunocompetence handicap hypothesis in male humans. *Scientific reports*, *8*(1), 7392. https://doi.org/10.1038/s41598-018-25694-0

O'Connor, M. C., & Paunonen, S. V. (2007). Big Five personality predictors of post-secondary academic performance. *Personality and Individual differences*, *43*(5), 971-990. https://doi.org/10.1016/j.paid.2007.03.017

O'Boyle Jr, E. H., Humphrey, R. H., Pollack, J. M., Hawver, T. H., & Story, P. A. (2011). The relation between emotional intelligence and job performance: A meta-analysis. *Journal of Organizational Behavior*, *32*(5), 788-818. https://doi.org/10.1002/job.714

Olkin, I., Dahabreh, I. J., & Trikalinos, T. A. (2012). GOSH–a graphical display of study heterogeneity. *Research Synthesis Methods, 3*(3), 214-223. https://doi.org/10.1002/jrsm.1053

Perrett, D. I., Burt, D. M., Penton-Voak, I. S., Lee, K. J., Rowland, D. A., & Edwards, R. (1999). Symmetry and human facial attractiveness. *Evolution and human behavior*, *20*(5), 295-307. https://doi.org/10.1016/S1090-5138(99)00014-8

Perrett, D. I., Lee, K. J., Penton-Voak, I., Rowland, D., Yoshikawa, S., Burt, D. M., ... & Akamatsu, S. (1998). Effects of sexual dimorphism on facial attractiveness. *Nature*, *394*(6696), 884-887. https://doi.org/10.1038/29772

Pesta, B. J., Kirkegaard, E. O., te Nijenhuis, J., Lasker, J., & Fuerst, J. G. (2020). Racial and ethnic group differences in the heritability of intelligence: A systematic review and meta-analysis. Intelligence, 78, 101408. https://doi.org/10.1016/j.intell.2019.101408

Petersen, J. L., & Hyde, J. S. (2010). A meta-analytic review of research on gender differences in sexuality, 1993–2007. *Psychological bulletin*, *136*(1), 21. https://doi.org/10.1037/a0017504

Petersen, J., & Hyde, J. S. (2010). Gender differences in sexuality. *Handbook of Gender Research in Psychology: Volume 1: Gender Research in General and Experimental Psychology*, 471-491. https://doi.org/10.1007/978-1-4419-1465-1_23

Petrill, S. A., Lipton, P. A., Hewitt, J. K., Plomin, R., Cherny, S. S., Corley, R., & DeFries, J. C. (2004). Genetic and environmental contributions to general cognitive ability through the first 16 years of life. *Developmental psychology*, *40*(5), 805. https://doi.org/10.1037/0012-1649.40.5.805

Pietschnig, J., & Voracek, M. (2015). One century of global IQ gains: A formal meta-analysis of the Flynn effect (1909–2013). *Perspectives on Psychological Science*, *10*(3), 282-306. https://doi.org/10.1177/1745691615577701

Pietschnig, J., Siegel, M., Eder, J. S. N., & Gittler, G. (2019). Effect declines are systematic, strong, and ubiquitous: A meta-meta-analysis of the decline effect in intelligence research. *Frontiers in psychology*, *10*, 2874. https://doi.org/10.3389/fpsyg.2019.02874

Plomin, R., & Deary, I. J. (2015). Genetics and intelligence differences: five special findings. *Molecular psychiatry*, *20*(1), 98-108. https://doi.org/10.1038/mp.2014.105

Plomin, R., & DeFries, J. C. (1980). Genetics and intelligence: Recent data. *Intelligence*, *4*(1), 15-24. https://doi.org/10.1016/0160-2896(80)90003-3

Plomin, R., & Spinath, F. M. (2004). Intelligence: genetics, genes, and genomics. *Journal of personality and social psychology*, *86*(1), 112. https://doi.org/10.1037/0022-3514.86.1.112

Pound, N., Penton-Voak, I. S., & Surridge, A. K. (2009). Testosterone responses to competition in men are related to facial masculinity. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1654), 153-159. https://doi.org/10.1098/rspb.2008.0990

Pound, N., Penton-Voak, I. S., & Surridge, A. K. (2009). Testosterone responses to competition in men are related to facial masculinity. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1654), 153-159. https://doi.org/10.1098/rspb.2008.0990

Procopio, F., Zhou, Q., Wang, Z., Gidziela, A., Rimfeld, K., Malanchini, M., & Plomin, R. (2022). The genetics of specific cognitive abilities. *Intelligence*, *95*, 101689. https://doi.org/10.1016/j.intell.2022.101689

Prokosch, M. D., Yeo, R. A., & Miller, G. F. (2005). Intelligence tests with higher g-loadings show higher correlations with body symmetry: Evidence for a general fitness factor mediated by developmental stability. *Intelligence*, *33*(2), 203-213. https://doi.org/10.1016/j.intell.2004.07.007

Pullar, J. M., Carr, A. C., & Vissers, M. (2017). The roles of vitamin C in skin health. *Nutrients*, *9*(8), 866. https://doi.org/10.3390/nu9080866

Purkey, A. (2011). The Heritability of Facial Attractiveness. https://scholar.colorado.edu/concern/undergraduate_honors_theses/6t053g52q

Puts, D. A., Bailey, D. H., Cárdenas, R. A., Burriss, R. P., Welling, L. L., Wheatley, J. R., & Dawood, K. (2013). Women's attractiveness changes with estradiol and progesterone across the ovulatory cycle. *Hormones and behavior*, *63*(1), 13-19. https://doi.org/10.1016/j.yhbeh.2012.11.007

Quinlan, R. J. (2008). Human pair-bonds: Evolutionary functions, ecological variation, and adaptive development. *Evolutionary Anthropology: Issues, News, and Reviews*, *17*(5), 227-238. https://doi.org/10.1002/evan.20191

R Core Team. (2022). R: A Language and Environment for Statistical Computing. R Core Team. Retrieved from https://www.R-project.org/

Rantala, M. J., Moore, F. R., Skrinda, I., Krama, T., Kivleniece, I., Kecko, S., & Krams, I. (2012). Evidence for the stress-linked immunocompetence handicap hypothesis in humans. *Nature communications*, *3*(1), 694. https://doi.org/10.1038/ncomms1696

Rantala, M. J., Moore, F. R., Skrinda, I., Krama, T., Kivleniece, I., Kecko, S., & Krams, I. (2012). Evidence for the stress-linked immunocompetence handicap hypothesis in humans. *Nature communications*, *3*(1), 694. https://doi.org/10.1038/ncomms1696

Regan, P. C., Levin, L., Sprecher, S., Christopher, F. S., & Gate, R. (2000). Partner preferences: What characteristics do men and women desire in their short-term sexual and long-term romantic partners?. *Journal of Psychology & Human Sexuality*, *12*(3), 1-21. https://doi.org/10.1300/J056v12n03_01

Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annu. Rev. Psychol.*, *57*, 199-226. https://doi.org/10.1146/annurev.psych.57.102904.190208

Rhodes, G., & Tremewan, T. (1996). Averageness, exaggeration, and facial attractiveness. *Psychological science*, *7*(2), 105-110. https://doi.org/10.1111/j.1467-9280.1996.tb00338

Rhodes, G., Proffitt, F., Grady, J. M., & Sumich, A. (1998). Facial symmetry and the perception of beauty. Psychonomic Bulletin & Review, 5, 659-669. https://doi.org/10.3758/BF03208842

Rhodes, G., Yoshikawa, S., Palermo, R., Simmons, L. W., Peters, M., Lee, K., Halberstadt, J., & Crawford, J. R. (2007). Perceived Health Contributes to the Attractiveness of Facial

Symmetry, Averageness, and Sexual Dimorphism. *Perception*, *36*(8), 1244–1252. https://doi.org/10.1068/p5712

Rhodes, G., Zebrowitz, L. A., Clark, A., Kalick, S. M., Hightower, A., & McKay, R. (2001). Do facial averageness and symmetry signal health?. *Evolution and human behavior*, *22*(1), 31-46. https://doi.org /10.1016/S1090-5138(00)00060-X

Roberts, M. L., Buchanan, K. L., & Evans, M. R. (2004). Testing the immunocompetence handicap hypothesis: a review of the evidence. *Animal behaviour*, *68*(2), 227-239. https://doi.org/10.1016/j.anbehav.2004.05.001

Roberts, S. C., Havlicek, J., Flegr, J., Hruskova, M., Little, A. C., Jones, B. C., ... & Petrie, M. (2004). Female facial attractiveness increases during the fertile phase of the menstrual cycle. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *271*(suppl_5), S270-S272. https://doi.org/10.1098/rsbl.2004.0174

Robinson, M. R., Kleinman, A., Graff, M., Vinkhuyzen, A. A., Couper, D., Miller, M. B., ... & Visscher, P. M. (2017). Genetic evidence of assortative mating in humans. *Nature Human Behaviour*, *1*(1), 0016. https://doi.org/10.1038/s41562-016-0016

Rosenblat, T. S. (2008). The beauty premium: Physical attractiveness and gender in dictator games. *Negotiation Journal*, *24*(4), 465-481. https://doi.org/10.1111/j.1571-9979.2008.00198.x

Ross, C. T., Borgerhoff Mulder, M., Oh, S. Y., Bowles, S., Beheim, B., Bunce, J., ... & Ziker, J. (2018). Greater wealth inequality, less polygyny: rethinking the polygyny threshold model. *Journal of the Royal Society interface*, *15*(144), 20180035. https://doi.org/10.1098/rsif.2018.0035

Rowe, D. C., Clapp, M., & Wallis, J. (1989). Physical attractiveness and the personality resemblance of identical twins. *Behavior Genetics, 17*, 191–201. https://doi.org/10.1007/BF01065997

Ryosuke, O., Yoshie, S., & Hiromi, A. (2021). The association between activity levels and skin moisturising function in adults. *Dermatology Reports*, *13*(1). https://doi.org/10.4081/dr.2021.8811

Sagebin Bordini, G., & Sperb, T. M. (2013). Sexual double standard: A review of the literature between 2001 and 2010. *Sexuality & Culture*, *17*, 686-704. https://doi.org/10.1007/s12119-012-9163-0

Savage, J. E., Jansen, P. R., Stringer, S., Watanabe, K., Bryois, J., De Leeuw, C. A., ... & Posthuma, D. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature genetics*, *50*(7), 912-919. https://doi.org/10.1038/s41588-018-0152-6

Scheib, J. E., Gangestad, S. W., & Thornhill, R. (1999). Facial attractiveness, symmetry and cues of good genes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *266*(1431), 1913-1917. https://doi.org/10.1098/rspb.1999.0866

Scheib, J. E., Gangestad, S. W., & Thornhill, R. (1999). Facial attractiveness, symmetry and cues of good genes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *266*(1431), 1913-1917. https://doi.org/10.1098/rspb.1999.0866

Schmitt, D. P. (2005). Sociosexuality from Argentina to Zimbabwe: A 48-nation study of sex, culture, and strategies of human mating. *Behavioral and Brain sciences*, *28*(2), 247-275. https://doi.org/10.1017/S0140525X05000051

Schnack, H. G., Van Haren, N. E., Brouwer, R. M., Evans, A., Durston, S., Boomsma, D. I., ... & Hulshoff Pol, H. E. (2015). Changes in thickness and surface area of the human cortex and their relationship with intelligence. *Cerebral cortex*, *25*(6), 1608-1617. https://doi.org/10.1093/cercor/bht357

Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence.

*Scholz, J. K., & Sicinski, K. (2015). Facial Attractiveness and Lifetime Earnings: Evidence from a Cohort Study. The Review of Economics and Statistics, 97(1), 14–28. https://doi.org/10.1162/REST_a_00435

Schooler, J. (2011). Unpublished results hide the decline effect. *Nature*, *470*(7335), 437-437. https://doi.org/10.1038/470437a

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). p-curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science*, *9*(6), 666-681. https://doi.org/10.1177/174569161455398

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2020). Specification curve analysis. Nature Human Behaviour, 4(11), 1208-1214. https://doi.org/10.1038/s41562-020-0912-z

Singh, D. (2004). Mating strategies of young women: Role of physical attractiveness. *Journal of sex Research*, *41*(1), 43-54. https://doi.org/10.1080/00224490409552212

Smith, M. J. L., Deady, D. K., Moore, F. R., Jones, B. C., Cornwell, R. E., Stirrat, M., Lawson, J. R., Feinberg, D. R., & Perrett, D. I. (2012). Maternal tendencies in women are associated with estrogen levels and facial femininity. *Hormones and Behavior*, *61*(1), 12-16. https://doi.org/10.1016/j.yhbeh.2011.09.005

Smith, M. J. L., Perrett, D. I., Jones, B. C., Cornwell, R. E., Moore, F. R., Feinberg, D. R., Boothroyd, L. G., Durrani, S. L., Stirrat, M. R., Whiten, S., Pitman, R. M., & Hillier,

S. G. (2006). Facial appearance is a cue to oestrogen levels in women. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1583), 135-140. https://doi.org/10.1098/rspb.2005.3296

Sniekers, S., Stringer, S., Watanabe, K., Jansen, P. R., Coleman, J. R., Krapohl, E., ... & Posthuma, D. (2017). Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nature genetics*, *49*(7), 1107-1112. https://doi.org/10.1038/ng.3869

Sprecher, S., Treger, S., & Sakaluk, J. K. (2013). Premarital sexual standards and sociosexuality: Gender, ethnicity, and cohort differences. *Archives of sexual Behavior*, *42*, 1395-1405. https://doi.org/10.1007/s10508-013-0145-6

Stern, Y. (2002). What is cognitive reserve? Theory and research application of the reserve concept. *Journal of the international neuropsychological society*, *8*(3), 448-460. https://doi.org/10.1017/S1355617702813248

Sternberg, R. J. (2019). A theory of adaptive intelligence and its relation to general intelligence. *Journal of Intelligence*, *7*(4), 23. https://doi.org/10.3390/jintelligence7040023

Sterne, J. A., & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. *Publication bias in meta-analysis: Prevention, assessment and adjustments*, 99-110. https://doi.org/10.1002/0470870168

Stewart-Williams, S., & Thomas, A. G. (2013). The ape that thought it was a peacock: Does evolutionary psychology exaggerate human sex differences?. *Psychological Inquiry*, *24*(3), 137-168. https://doi.org/10.1080/1047840X.2013.804899

Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence*, *35*(5), 401-426. https://doi.org/10.1016/j.intell.2006.09.004

Strenze, T. (2015). Intelligence and success. *Handbook of intelligence: Evolutionary theory, historical perspective, and current concepts*, 405-413. https://doi.org/10.1007/978-1-4939-1562-0_25

Švegar, D. (2016). What does facial symmetry reveal about health and personality?. *Polish Psychological Bulletin*, *47*(3), 356-365.

Thornhill, R., & Gangestad, S. W. (1993). Human facial beauty: Averageness, symmetry, and parasite resistance. *Human nature*, *4*, 237-269. https://doi.org/10.1007/BF02692201

Thornhill, R., & Gangestad, S. W. (1999). Facial attractiveness. *Trends in cognitive sciences*, *3*(12), 452-460. https://doi.org/10.1016/S1364-6613(99)01403-5

Thornhill, R., & Gangestad, S. W. (2006). Facial sexual dimorphism, developmental stability, and susceptibility to disease in men and women. *Evolution and Human Behavior*, *27*(2), 131-144. https://doi.org/10.1016/j.evolhumbehav.2005.06.001

Thornhill, R., & Møller, A. P. (1997). Developmental stability, disease and medicine. *Biological Reviews*, *72*(4), 497-548. https://doi.org/10.1017/S0006323197005082

Todorov, A., Dotsch, R., Porter, J. M., Oosterhof, N. N., & Falvello, V. B. (2013). Validation of data-driven computational models of social perception of faces. *Emotion*, *13*(4), 724. https://doi.org/10.1037/a0032335

Trigunaite, A., Dimo, J., & Jørgensen, T. N. (2015). Suppressive effects of androgens on the immune system. *Cellular immunology*, *294*(2), 87-94. https://doi.org/10.1016/j.cellimm.2015.02.004

Trivers, R. (1972). Parental investment and sexual selection. In B. Campbell (Ed.), *Sexual selection and the descent of man* (pp. 136-179). Chicago: Aldine.

Trujillo, L. T., Jankowitsch, J. M., & Langlois, J. H. (2014). Beauty is in the ease of the beholding: A neurophysiological test of the averageness theory of facial attractiveness. *Cognitive, Affective, & Behavioral Neuroscience*, *14*, 1061-1076. https://doi.org/10.3758/s13415-013-0230-2

Tsagkrasoulis, D., Hysi, P., Spector, T., & Montana, G. (2017). Heritability maps of human face morphology through large-scale automated three-dimensional phenotyping. *Scientific reports*, *7*(1), 45885. https://doi.org/10.1038/srep45885

Tsagkrasoulis, D., Hysi, P., Spector, T., & Montana, G. (2017). Heritability maps of human face morphology through large-scale automated three-dimensional phenotyping. *Scientific reports*, *7*(1), 45885. https://doi.org/10.1038/srep45885

Tucker-Drob, E. M., Rhemtulla, M., Harden, K. P., Turkheimer, E., & Fask, D. (2011). Emergence of a gene× socioeconomic status interaction on infant mental ability between 10 months and 2 years. *Psychological science*, *22*(1), 125-133. https://doi.org/10.1177/0956797610392926

Turkheimer, E., Haley, A., Waldron, M., d'Onofrio, B., & Gottesman, I. I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological science*, *14*(6), 623-628. https://doi.org/10.1046/j.0956-7976.2003.psci_1475.x

Twenge, J. M., Sherman, R. A., & Wells, B. E. (2015). Changes in American adults' sexual behavior and attitudes, 1972–2012. *Archives of sexual behavior*, *44*, 2273-2285. https://doi.org/10.1007/s10508-015-0540-2

Valentine, T., Darling, S., & Donnelly, M. (2004). Why are average faces attractive? The effect of view and averageness on the attractiveness of female faces. *Psychonomic Bulletin & Review*, *11*, 482-487. https://doi.org/10.3758/BF03196599

van Aert, R. C., & van Assen, M. A. (2021). Correcting for publication bias in a meta-analysis with the *p*-uniform* method. https://doi.org/10.31222/osf.io/zqjr9

van Assen, M. A., van Aert, R., & Wicherts, J. M. (2015). Meta-analysis using effect size distributions of only statistically significant studies. *Psychological methods*, *20*(3), 293. https://doi.org/10.1037/met0000025

Voracek, M., Kossmeier, M., & Tran, U. S. (2019). Which data to meta-analyze, and how? *Zeitschrift fur Psychologie*. https://doi.org/10.1027/2151-2604/a000357

Watkins, C. D., Jones, B. C., & DeBruine, L. M. (2010). Individual differences in dominance perception: Dominant men are less sensitive to facial cues of male dominance. *Personality and Individual Differences*, *49*(8), 967-971. https://doi.org/10.1016/j.paid.2010.08.006

White, D. R., & Burton, M. L. (1988). Causes of polygyny: Ecology, economy, kinship, and warfare. *American Anthropologist*, *90*(4), 871-887. https://doi.org/10.1525/aa.1988.90.4.02a00060

Yu, C. C., Furukawa, M., Kobayashi, K., Shikishima, C., Cha, P. C., Sese, J., ... & Toda, T. (2012). Genome-wide DNA methylation and gene expression analyses of monozygotic twins discordant for intelligence levels. https://doi.org/10.1371/journal.pone.0047081

Zaboski II, B. A., Kranzler, J. H., & Gage, N. A. (2018). Meta-analysis of the relationship between academic achievement and broad abilities of the Cattell-Horn-Carroll theory. *Journal of school psychology*, *71*, 42-56. https://doi.org/10.1016/j.jsp.2018.10.001

Zaidel, D. W., & Hessamian, M. (2010). Asymmetry and symmetry in the beauty of human faces. *Symmetry*, *2*(1), 136-149. https://doi.org/10.3390/sym2010136

Zaidel, D. W., Aarde, S. M., & Baig, K. (2005). Appearance of symmetry, beauty, and health in human faces. *Brain and cognition*, *57*(3), 261-263. https://doi.org/10.1016/j.bandc.2004.08.056

Zebrowitz, L. A., Collins, M. A., & Dutta, R. (1998). The relationship between appearance and personality across the life span. *Personality and social psychology bulletin*, *24*(7), 736-749. https://doi.org/10.1177/0146167298247006

Zebrowitz, L. A., Franklin Jr, R. G., Hillman, S., & Boc, H. (2013). Older and younger adults' first impressions from faces: similar in agreement but different in positivity. *Psychology and aging*, *28*(1), 202. https://doi.org/10.1037/a0030927

*Zebrowitz, L. A., Hall, J. A., Murphy, N. A., & Rhodes, G. (2002). Looking smart and looking good: Facial cues to intelligence and their origins. Personality and Social Psychology Bulletin, 28(2), 238–249. https://doi.org/10.1177/0146167202282009

Zebrowitz, L. A., Olson, K., & Hoffman, K. (1993). Stability of babyfaceness and attractiveness across the life span. *Journal of personality and social psychology*, *64*(3), 453. https://doi.org/10.1037/0022-3514.64.3.453

Żelaźniewicz, A., Bielawski, T., Nowak, J., & Pawłowski, B. (2019). Body symmetry and reproductive hormone levels in women. *Women & Health*, *59*(4), 391-405. https://doi.org/10.1080/03630242.2018.1492499

Zerjal, T., Xue, Y., Bertorelle, G., Wells, R. S., Bao, W., Zhu, S., ... & Tyler-Smith, C. (2003). The genetic legacy of the Mongols. *The American Journal of Human Genetics*, *72*(3), 717-721. https://doi.org/10.1086/367774

Zheng, R., Ren, D., Xie, C., Pan, J., & Zhou, G. (2021). Normality mediates the effect of symmetry on facial attractiveness. *Acta Psychologica*, *217*, 103311. https://doi.org/10.1016/j.actpsy.2021.103311

List of Figures

Figure 1. *Flow-chart of study inclusion* 22
Figure 2. *Forest plots. Panel A displays the full dataset (k = 14), panel B shows the trimmed dataset without the outlier dataset (k = 13).* 28
Figure 3. *Contour-enhanced funnel plots with imputed trim-and-fill values and Egger's regression line (left, Fig. 3A, 3C) and power-enhanced funnel plots (right, Fig. 3B, 3D). The top line shows funnel plots of the full dataset (k = 14), the bottom line displays the trimmed dataset (k = 13). Study power ranges from 0 to 10 percent (dark red) to 80 to 90 percent (dark green) in 10 percent increments. Median average power was 43.8% for the full analysis (Fig. 3B), and 26.3% for the trimmed analysis (Fig. 3D).* 34
Figure 4. *p-Curve based on eight statistically significant results (p < .05), of which six are p < .025. Six additional records were excluded from p-curve because they were p > .05. The blue line represents the distribution of observed p-values. The red dashed line represents the expected distribution under the null hypothesis of no effect. The green dashed line displays the condition of a true effect and underpowered studies (33% power).* 35
Figure 5. *Specification Curve Analysis. The bottom panel represents "which" and "how" factors. The middle panel represents the number of samples within the subsets. The top panel displays the estimated summary effects with respective 95% of confidence intervals.* 37
Figure 6. *GOSH-Plots. Each dot represents the summary effect of a random subset of studies. Panel A displays the full data (k = 14), which resulted in 16,383 combinations. Panel B shows the trimmed data (k = 13), yielding 8,191 different combinations. Subset estimations including at least one of the studies with larger effect sizes are visualized in red. Distributional densities are shown on the top (summary effects) and on the right side (I^2 values) of the plot.* 38

List of Tables

Table 1. Details of included studies. 24

Table 2. Random-effects estimates of overall data and subgroup analyses. 29

Table 3. *Numeric outcomes of moderator analyses.* 29

Table 4. Numeric outcomes of linear-precision-weighted meta-regression. 30

Table 5. Leave-one-out analyses. 31

Table S1. Prisma checklist. 76

Table S2. Quality of included primary studies assessed with the Newcastle-Ottawa Scale.... 80

Table S3. Study characteristics of excluded studies. 81

Table S4. Numeric outcomes of publication bias detection methods. 96

Appendix A

Abstracts in English and German

Abstract (English)

The correlation between physical attractiveness and intelligence has long intrigued evolutionary theorists. One prominent explanation is based on cross-trait assortative mating (xAM), where individuals with high levels of one desirable trait tend to pair with those who have high levels of another desirable trait. Two large meta-analyses have shown a small positive correlation between physical attractiveness and intelligence. However, prior meta-analyses have incorporated studies utilizing measures such as grade point averages, school performance, and academic achievement to evaluate intelligence, potentially biasing the outcomes. The present meta-analysis is the first to exclusively analyze studies that used psychometric IQ-tests $n = 9$ ($k = 14$, $N = 50{,}156$). Random-effects analyses showed a small yet robust positive association between physical attractiveness and intelligence $r = 0.13$ ($p = .0005$, 95% CI [0.06, 0.20]). Even after removing an outlier study that showed an exceptionally high correlation, the association between physical attractiveness and intelligence remained robust albeit slightly smaller $r = 0.10$ ($p = .0004$, 95% CI [0.04, 0.15]). No strong evidence was found for publication bias. Moreover, combinatorial analyses provided further support for the robustness of the results. Overall, the results indicate a small and robust association between physical attractiveness and intelligence.

Keywords: Intelligence, physical attractiveness, cross-trait assortative mating, meta-analysis, multiverse-analysis.

Abstract (German)

Die Korrelation zwischen physischer Attraktivität und Intelligenz fasziniert Evolutionstheoretiker seit langem. Eine der populärsten Erklärungen basiert auf dem Prinzip von cross-trait assortative mating (xAM), bei der sich Individuen mit hohen Ausprägungen eines begehrenswerten Merkmals tendenziell mit Personen paaren, die hohe Ausprägungen eines anderen begehrenswerten Merkmals besitzen. Zwei umfassende Meta-Analysen haben einen kleinen positiven Zusammenhang zwischen Attraktivität und Intelligenz gezeigt. Allerdings wurden in den bisherigen Meta-Analysen Studien analysiert, die zur Bewertung der Intelligenz unter anderem Notendurchschnitte, Schulleistungen und akademische Erfolge verwendeten, was die Ergebnisse potenziell verzerrt. Die vorliegende Meta-Analyse ist die erste, die ausschließlich Studien analysiert, in denen Intelligenz mit psychometrischen IQ-Tests gemessen wurde n = 9 (k = 14, N = 50.156). Random-effects models zeigten einen kleinen, robusten positiven Zusammenhang zwischen körperlicher Attraktivität und Intelligenz r = 0.13 (p = .0005, 95% CI [0.06, 0.20]). Selbst nach Exklusion einer Ausreißerstudie, welche eine außergewöhnlich hohe Korrelation aufwies, blieb der Zusammenhang zwischen körperlicher Attraktivität und Intelligenz stabil, wenn auch etwas kleiner r = 0.10 (p = .0004, 95% CI [0.04, 0.15]). Es wurden keine eindeutigen Hinweise auf einen Publikationsbias gefunden. Darüber hinaus lieferten kombinatorische Analysen einen weiteren Beleg für die Robustheit der Ergebnisse. Insgesamt deuten die Ergebnisse auf einen kleinen und robusten Zusammenhang zwischen Attraktivität und Intelligenz hin.

Schlüsselwörter: Intelligenz, physische Attraktivität, cross-trait assortative mating, Meta-Analyse, Multiversum-Analyse.

Appendix B

Deviations from preregistration protocol

1. Hypotheses that were not investigated due to insufficient/missing statistical information:
Hypothesis 2: *There is a positive association between childhood attractiveness and adult IQ.*
Hypothesis 4: *Participants' sex has no significant effect on the association between attractiveness and IQ.*

2. Moderators: only age (children vs. mixed vs. adults) and attractiveness assessment (live vs. photographs) were assessed.
Additional moderators that were intended to be investigated: intelligence assessment (e.g., Stanford vs. Wechsler), status of publication (published vs. grey literature).

3. Multiverse meta-analysis/specification curve analysis: "which" attractiveness factors that were excluded: men only vs. women only, intelligence assessment, attractiveness assessment, publication status.

4. Literature search in Scopus: results limited to 'Psychology'.

5. Analyses were performed twice: once on the full dataset and a second time with an outlier sample removed (Kanazawa, 2011 – Group A).

Appendix C

PRISMA checklist

Table S1. *Prisma checklist.*

Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	Cover page
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	1
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	2
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	16
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	17
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	17, 11, 119
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	17
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	17, 18
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	17, 18
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	17, 18
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	17, 18
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	18 - 21
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	18 - 21
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	17
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	17

Section and Topic	Item #	Checklist item	Location where item is reported
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	18 - 21
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	18 - 21
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	18
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	20, 21
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	18, 19
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	18 - 21
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	22, 23
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	81 - 88
Study characteristics	17	Cite each included study and present its characteristics.	24, 50 - 70
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	79
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	26, 28, 29
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	23
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	18, 19
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	18, 19, 20
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	22-26
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	22
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	18-26
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	39 - 45
	23b	Discuss any limitations of the evidence included in the review.	45, 46
	23c	Discuss any limitations of the review processes used.	45, 46

Section and Topic	Item #	Checklist item	Location where item is reported
	23d	Discuss implications of the results for practice, policy, and future research.	46, 47
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	16, 17
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	16, 17
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	17, 75
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	-
Competing interests	26	Declare any competing interests of review authors.	-
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	16, 17

Appendix D

Newcastle-Ottawa Scale

Selection: (Maximum 5 stars, adapted 3 stars)

1. Representativeness of the sample:
    a. Truly representative of the average in the target population. * (all subjects or random sampling)
    b. Somewhat representative of the average in the target group. * (non-random sampling)
    c. Selected group of users/convenience sample.
    d. No description of the derivation of the included subjects.

2. Sample size:
    a. Justified and satisfactory (including sample size calculation). *
    b. Not justified.
    c. No information provided

3. Non-respondents:
    a. Proportion of target sample recruited attains pre-specified target or basic summary of non-respondent characteristics in sampling frame recorded. *
    b. Unsatisfactory recruitment rate, no summary data on non-respondents.
    c. No information provided

4. Ascertainment of the exposure (risk factor): excluded
    a. Vaccine records/vaccine registry/clinic registers/hospital records only. **
    b. Parental or personal recall and vaccine/hospital records. *
    c. Parental/personal recall only.

Comparability: (Maximum 2 stars)

1. Comparability of subjects in different outcome groups on the basis of design or analysis. Confounding factors controlled.
    a. Data/ results adjusted for relevant predictors/risk factors/confounders e.g. age, sex, time since vaccination, etc. **
    b. Data/results not adjusted for all relevant confounders/risk factors/information not provided.

Outcome: (Maximum 3 stars)

1. Assessment of the outcome:
    a. Independent blind assessment. **
    b. Record linkage. **
    c. Self report. *
    d. No description.

2. Statistical test:
    a. Statistical test used to analyse the data clearly described, appropriate and measures of association presented including confidence intervals and probability level (p value). *

b. Statistical test not appropriate, not described or incomplete.

Table S2. *Quality of included primary studies assessed with the Newcastle-Ottawa Scale.*

Study	Selection				Comparability	Outcome		Total (8)
	1.	2.	3.	4.		1.	2.	
Kanazawa (2011)	*	*	*	-	**	**	*	8
Colladon et al. (2018)	*	*	0	-	0	**	*	5
Lee et al. (2017)	*	*	*	-	0	**	*	6
Judge et al. (2009)	*	*	0	-	**	**	*	7
Kleisner et al. (2014)	*	c	*	-	0	**	*	5
Zebrowitz et al. (2002)	*	*	*	-	0	**	*	6
Gangestad et al. (2010)	*	c	*	-	0	**	*	5
Härtel et al. (2023)	*	*	*	-	0	**	*	6
Scholz & Sicinski (2015)	*	*	0	-	0	**	*	5

Herzog, R., Alvarez-Pasquin, M., Diaz, C., Del Barrio, J., Estrada, J., & Gil, A. (2013). Newcastle-Ottawa Scale adapted for cross-sectional studies. *BMC Public Health*, *13*, 154.

Appendix E

Table S3. *Study characteristics of excluded studies.*

Reference	Sample size	Sample type	Country	Conclusion	Publication	Exclusion reason	Journal
Wheeler (1986)	24	Children	United States	Study investigated a physical attractiveness bias in scoring the Verbal Scoring of the Wechsler Intelligence Scale for Children (Revised) (WISC-R). No evidence for a physical attractiveness stereotype was found.	Peer-reviewed article	- Measured verbal IQ only. - No information provided of raw attractiveness scores. - No contact information available. - Statistical information missing to calculate effect size.	Journal of General Psychology
Langlois et al. (2000)	3,853	Mixed	-	Intelligence and physical attractiveness are barely correlated.	Meta-analysis	- Other metrics to measure intelligence were used (i.e., American College Test (ACT), Iowa Test of Basic Skills, vocabulary, GPA).	Psychological Bulletin
Feingold (1992)	15,205	Mixed	-	Trivial relationship between physical attractiveness and mental ability. Attractive people were found to be less lonely, less socially anxious, more popular, more socially skilled, and more sexually experienced than unattractive people.	Meta-analysis	- Intelligence was defined as performance on standardized tests (i.e., Scholastic Aptitude Test (SAT), American College Test (ACT). - Several included studies also used self-reported SAT scores.	Psychological Bulletin
Tucker & O'Grady (1991)	128	Adults	United States	The interaction between male and female intelligence was found to be the strongest	Peer-reviewed article	- Used GPA and SAT scores as intelligence measurements.	The Journal of Social Psychology

				predictor of marital satisfaction.			
Sparacino & Hansell (1979)	$N_1 = 120$ $N_2 = 50$	Adults	United States	Less attractive individuals surpassed attractive subjects in academic achievement.	Peer-reviewed article	- Used ACT scores as intelligence assessment. - No contact information available.	Journal of Personality
Feingold (1982)	150	Adults	-	Nonsignificant correlation between physical attractiveness and IQ.	Peer-reviewed article	- Data requested, data unavailable (first two pages available).	The Journal of Social Psychology
White et al. (2006)	$N_1 = 16$ $N_2 = 24$	Adults	United Kingdom	Individuals with Asperger Syndrome were generally competent in making judgments of trustworthiness, attractiveness social status and age, with only a trend towards potential difficulty in judging attractiveness based on the same sex, which was attributed to challenges in mentalizing.	Peer-reviewed article	- Investigated patients with Asperger Syndrome (N_1) and healthy controls (N_2). - Missing attractiveness ratings of adults with Asperger Syndrome.	Brain and Cognition
Mitchem et al. (2015)	$N_1 = 399$ $N_2 = 1354$	Mixed	United States, Australia	No evidence for a phenotypic correlation between facial attractiveness and IQ.	Peer-reviewed article	- Identical data already included in another study (see Lee et al., 2017) with similar results.	Evolution and Human Behavior
Hollingworth (1935)	40	Mixed	United States	Highly intelligent individuals were perceived as more attractive compared to faces from the ordinary intelligence group.	Peer-reviewed article	- No Pearson correlation between physical attractiveness and intelligence reported.	The Pedagogical Seminary and Journal of Genetic Psychology
Eagly et al. (1991)	76 studies (exact	Mixed	United States,	No strong/consistent relationship between physical	Meta-analysis	- Authors did not specify intelligence assessment.	Psychological Bulletin

	number not specified)		Canada	attractiveness and intelligence.			
Benzeval et al. (2013)	1,515	Mixed	United Kingdom/Scotland	More phyiscally attractive people at age 15 had higher socioeconomic positions at age 36.	Peer-reviewed article	- No valid IQ-test at age 15.	Plos One
Furnham & Robinson (2023)	2,020	Adults	Europe (not specified)	Investigated correlates of self-assessed optimism. Participants rated to what extent they were an optimist, as well as self-ratings of four variables (e.g., attractiveness, intelligence).	Peer-reviewed article	- Only self-ratings of intelligence and physical attractiveness. - No metric IQ-test.	Current Research in Behavioral Sciences
Kanazawa & Reyniers (2009)	20,745	Adults	United States	Taller men have higher IQs. Association between physical attractiveness and IQ explained by assortative mating of tall men and beautiful women.	Peer-reviewed article	- No psychometric IQ-test; study deployed the Peabody Vocabulary Test (PPVT)	The American Journal of Psychology
Kanazawa & Still (2018)	20,745	Adults	United States	Positive correlation between physical attractiveness and IQ.	Peer-reviewed article	- No psychometric IQ-test; study deployed the Peabody Vocabulary Test (PPVT)	Journal of Business and Psychology
Driebe et al. (2021)	88	Adults	Germany	Intelligence can be detected via videos with above hance accuracy by members of the opposite sex.	Peer-reviewed article	- Information to calculate effect sizes was unavailable. - No author response.	Evolution and Human Behavior

Zebrowitz & Rhodes (2004)	$N_1 = 103$ $N_2 = 81$ $N_3 = 62$ $N_4 = 101$ $N_5 = 98$ $N_6 = 68$	Mixed	United States	Attractiveness was significantly correlated with IQ scores in childhood, puberty, and adulthood for faces in the lower half of the attractiveness distribution, whereas for faces in the lower half of distribution, correlations were not significant.	Peer-reviewed article	- Identical data already included in another study (see Zebrowitz et al., 2002).	The Journal of Nonverbal Behavior
Sim et al. (2015)	161	Adults	United States	Attractiveness judgements are influenced by the evaluator's personal traits (such as attractiveness and IQ).	Peer-reviewed article	- Statistical information to calculate effect size was unavailable (attractiveness ratings from judges). - Only self-ratings of attractiveness available in study. - No author response.	Social Psychology
Olderbak et al. (2021a)	339	Adults	Germany	-	Peer-reviewed article	- Statistical information to calculate effect size was unavailable. - Subjects were German prisoners. - Test scores were taken from another study (Olderbak et al., 2021b) - no access to study. - No author response.	Journal of Intelligence
DeBrabander et al. (2019)	40	Adults	United States	Autistic as well as typically developing adults evaluated autistic adults less favorably (e.g., in attractiveness) than typically developing adults.	Peer-reviewed article	- Autistic test subjects and healthy controls. - Information to calculate effect size was unavailable. - No author response.	Autism in Adulthood
LaChapelle et al. (1999)	40	Adults	-	-	Peer-reviewed article	- No access to paper. - Subjects had intellectual disability. - No author response.	The Clinical Journal of Pain

Hofer et al. (2021)	180	Adults	Austria	-	Journal pre-proof	- Statistical information to calculate effect size was unavailable. - No author response.	Journal of Research in Personality
Costa & Maestripieri (2023)	231	Adults	-	Attractiveness was negatively correlated with body mass index and with height (only in males).	Peer-reviewed article	- No access to paper. - No author response.	Evolutionary Behavioral Sciences
Gao et al. (2017)	69	Adults	China	-	Peer-reviewed article	- Statistical information to calculate effect size was unavailable. - No author response.	Frontiers in Psychology

Note. N = sample size;

References Appendix E

Benzeval, M., Green, M. J., & Macintyre, S. (2013). Does perceived physical attractiveness in adolescence predict better socioeconomic position in adulthood? Evidence from 20 years of follow up in a population cohort study. *PLoS One*, *8*(5), e63975. https://doi.org/10.1371/journal.pone.0063975

Costa, M., & Maestripieri, D. (2023). Physical and psychosocial correlates of facial attractiveness. *Evolutionary Behavioral Sciences*. https://psycnet.apa.org/doi/10.1037/ebs0000331

Costa, M., & Maestripieri, D. (2023). Physical and psychosocial correlates of facial attractiveness. *Evolutionary Behavioral Sciences*. https://doi.org/10.1037/ebs0000331

DeBrabander, K. M., Morrison, K. E., Jones, D. R., Faso, D. J., Chmielewski, M., & Sasson, N. J. (2019). Do first impressions of autistic adults differ between autistic and nonautistic observers?. *Autism in Adulthood*, *1*(4), 250-257. https://doi.org/10.1089/aut.2019.0018

Driebe, J. C., Sidari, M. J., Dufner, M., Von der Heiden, J. M., Bürkner, P. C., Penke, L., Zietsch, B. P., & Arslan, R. C. (2021). Intelligence can be detected but is not found attractive in videos and live interactions. *Evolution and Human Behavior*, *42*(6), 507-516. https://doi.org/10.1016/j.evolhumbehav.2021.05.002

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but…: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological bulletin*, *110*(1), 109. https://doi.org/10.1037/0033-2909.110.1.109

Feingold, A. (1982). Physical attractiveness and intelligence. *The Journal of Social Psychology*. https://doi.org/10.1080/00224545.1982.9922810

Feingold, A. (1992). Good-looking people are not what we think. *Psychological bulletin*, *111*(2), 304. https://doi.org /10.1037/0033-2909.111.2.304

Furnham, A., & Robinson, C. (2023). Correlates of self-assessed optimism. *Current Research in Behavioral Sciences*, *4*, 100089. https://doi.org/10.1016/j.crbeha.2022.100089

Gao, Z., Yang, Q., Ma, X., Becker, B., Li, K., Zhou, F., & Kendrick, K. M. (2017). Men who compliment a Woman's appearance using metaphorical language: Associations with creativity, masculinity, intelligence and attractiveness. *Frontiers in psychology*, *8*, 2185. https://doi.org/10.3389/fpsyg.2017.02185

Hofer, G., Burkart, R., Langmann, L., & Neubauer, A. C. (2021). What you see is what you want to get: Perceived abilities outperform objective test performance in predicting mate appeal in speed dating. *Journal of Research in Personality*, *93*, 104113. https://doi.org/10.1016/j.jrp.2021.104113

Hollingworth, L. S. (1935). The comparative beauty of the faces of highly intelligent adolescents. *The Pedagogical Seminary and Journal of Genetic Psychology*, *47*(2), 268-281. https://doi.org /10.1080/08856559.1935.10534047

Kanazawa, S., & Reyniers, D. J. (2009). The role of height in the sex difference in intelligence. *The American journal of psychology*, *122*(4), 527-536. https://doi.org/10.2307/27784427

Kanazawa, S., & Still, M. C. (2018). Is there really a beauty premium or an ugliness penalty on earnings?. *Journal of Business and Psychology*, *33*(2), 249-262. https://doi.org/10.1007/s10869-017-9489-6

LaChapelle, D. L., Hadjistavropoulos, T., & Craig, K. D. (1999). Pain measurement in persons with intellectual disabilities. *The Clinical journal of pain*, *15*(1), 13–23. https://doi.org/10.1097/00002508-199903000-00004

Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological bulletin*, *126*(3), 390. https://doi.org/10.1037/0033-2909.126.3.390

Mitchem, D. G., Zietsch, B. P., Wright, M. J., Martin, N. G., Hewitt, J. K., & Keller, M. C. (2015). No relationship between intelligence and facial attractiveness in a large, genetically informative sample. *Evolution and Human Behavior*, *36*(3), 240-247. https://doi.org /10.1016/j.evolhumbehav.2014.11.009

Olderbak, S. G., Geiger, M., Hauser, N. C., Mokros, A., & Wilhelm, O. (2021b). Emotion expression abilities and psychopathy. *Personality Disorders: Theory, Research, and Treatment*, *12*(6), 546. https://doi.org/10.1037/per0000444

Olderbak, S., Bader, C., Hauser, N., & Kleitman, S. (2021a). Detection of psychopathic traits in emotional faces. *Journal of Intelligence*, *9*(2), 29. https://doi.org/10.3390/jintelligence9020029

Sim, S. Y. L., Saperia, J., Brown, J. A., & Bernieri, F. J. (2015). Judging attractiveness: Biases due to raters' own attractiveness and intelligence. *Cogent Psychology*, *2*(1), 996316. https://doi.org/10.1080/23311908.2014.996316

Sparacino, J., & Hansell, S. (1979). Physical attractiveness and academic performance: Beauty is not always talent 1. *Journal of Personality*, *47*(3), 449-469. https://doi.org/10.1111/j.1467-6494.1979.tb00626.x

Tucker, M. W., & O'Grady, K. E. (1991). Effects of physical attractiveness, intelligence, age at marriage, and cohabitation on the perception of marital satisfaction. *The Journal of Social Psychology*, *131*(2), 253-269. https://doi.org/10.1080/00224545.1991.9713848

Wheeler, Paula Theisler, "A Study of the Effect of a Child's Physical Attractiveness upon Verbal Scoring of the Wechsler Intelligence Scale for Children (Revised) and upon Personality Attributions" (1985). *All Graduate Theses and Dissertations, Spring 1920 to Summer 2023*. 5937. https://digitalcommons.usu.edu/etd/5937

White, S., Hill, E., Winston, J., & Frith, U. (2006). An islet of social ability in Asperger Syndrome: judging social attributes from faces. *Brain and cognition*, *61*(1), 69-77. https://doi.org/10.1016/j.bandc.2005.12.007

Zebrowitz, L. A., & Rhodes, G. (2004). Sensitivity to “bad genes” and the anomalous face overgeneralization effect: Cue validity, cue utilization, and accuracy in judging intelligence and health. *Journal of nonverbal behavior, 28*, 167-185. https://doi.org/10.1023/B:JONB.0000039648.30935.1b

Appendix F

Descriptions of IQ-tests of included studies

Stanford-Binet Intelligence Scale Fourth Edition (SB4; Youngstrom et al., 2003). The Stanford-Binet Intelligence Scale: Fourth Edition (SB4) is structured in three tiers. The first tier focuses on the Composite, a measure of general ability associated with Binet scales. In the second tier, SB4 introduces a new element by suggesting three group factors: Crystallized Abilities, Fluid-Analytic Abilities, and Short-Term Memory. The first two dimensions are rooted in the Cattell-Horn theory of intelligence (Cattell, 1940; Horn, 1968; Horn & Cattell, 1966). The third tier describes factors based on three facets of reasoning: Verbal Reasoning, Quantitative Reasoning, and Abstract Visual Reasoning. Reliability scores for the subtests range from .73 to .94. SB4 has garnered a substantial body of validity evidence (see Youngstrom et al., 2003, p. 229).

Wechsler Adult Intelligence Scale (WAIS, Wechsler, 1955), Wechsler Adult Intelligence Scale-Revised (WAIS-R, Wechsler, 1981). The Full-scale IQ score is composed of a Verbal IQ and a Performance IQ score. To assess these scores, several subtests are applied (the Verbal Comprehension Index, the Working Memory Index, the Perceptual Organization Index and the Processing Speed Index). The WAIS-R is designed to measure cognitive ability in adults through various subtests that assess verbal comprehension, perceptual organization, working memory, and processing speed. Reliability evidence for the WAIS-R shows internal consistency coefficients ranging from .93 to .98 for the Full-Scale IQ (Wechsler, 1981). Validity studies have demonstrated strong correlations with other established measures of intelligence (Ryan & Ward, 1999).

Raven Progressive Matrices Test (RPM) (Raven et al., 1962; Arthur & Day, 1994; Raven, 2003). The RPM is a non-verbal assessment commonly employed to gauge overall human intelligence and abstract reasoning. It is recognized as a non-verbal approximation of fluid intelligence (Bilker et al., 2012). The test involves a visual geometric pattern featuring a vacant segment, accompanied by six to eight options to complete the missing part. Generally, the RPM has good psychometric properties, including high reliability and validity and has been found to have correlations with other IQ tests. The RPM tests have been shown to have moderate to high correlations with other intelligence tests, such as the Stanford-Binet and Wechsler scales, with correlations ranging from 0.54 to 0.86. Furthermore, RPM scores are associated with school and occupational performance (Raven, 2003).

Intelligence Structure Test 2000R (IST 2000-R) (Amthauer, 2001). The IST 2000-R is a multidimensional paper-pencil IQ-test comprising of several subtests to measure fluid and crystalline intelligence. The IST 2000-R includes a wide range of subtests, but this description focuses specifically on the subtests used to determine logical reasoning, which are 'Analogies', 'Number Series', and 'Matrices'. These three subtests provide an overall score for logical reasoning. The Cronbach's Alpha values for the subtests 'Analogies' and 'Matrices' are notably low, at .74 and .72, respectively. However, the corresponding test-retest and parallel test indices are deemed positive (Petermann, 2014). Despite the lower internal consistency, the overall reliability and validity of the IST 2000-R are supported by these indices, making it a valuable tool for assessing multiple facets of intelligence.

Henmon-Nelson Test of Mental Ability (Henmon, 1958). Each tier comprises 90 multiple-choice questions presented in a spiral-omnibus format, which means that the questions are arranged in increasing order of difficulty and cover a wide range of content areas, including various verbal items, numerical items, and figure analogies (Lamke & Nelson, 1958). The spiral-omnibus format ensures that the test measures a broad spectrum of cognitive abilities. Watson et al. (1992) demonstrated that the Henmon-Nelson Test explained approximately 50% additional variance in WAIS-R Verbal and Full-Scale IQs, indicating its strong predictive validity in the context of assessing general intelligence.

The Multidimensional Aptitude Battery (MAB) (Jackson, 1984; Jackson 1998). The MAB is a group-administered intelligence test comprising of ten subtests intended to assess Verbal, Performance, and Full-Scale IQ (Jackson, 1998). Krieshok & Harrington (1985) mention a series of studies on reliability and validity of the MAB, revealing consistently robust psychometric characteristics in their findings. In one study with over 500 participants, internal consistency reliability ranged from .94 to .97. for the Verbal scale, .95 to .98 for the Performance scale, and .96 to .98 for the Full scale.

Peabody Picture Vocabulary Test (PPVT) (Dunn & Dunn, 1965). The PPVT is a verbally administered IQ test, which predominantly measures crystalline/verbal intelligence. It differs from traditional IQ-tests as it does not directly assess fluid intelligence. The test includes 204 picture items. For each round, four pictures are shown, while the examiner reads out words describing one of the pictures. The individual is then asked to point the picture that the word describes. Norms are based on 828 individuals (ages ranging from 19 to 40) (Dunn & Dunn, 1965). Reliability and validity are considered to be satisfactory, with alpha coefficients ranging from .92 to .98 and split-half reliability from .86 to .97. for both Form IIIA and IIIB (Williams, 1999).

References Appendix F

Amthauer, R., Brocke, B., Liepmann, D., & Beauducel, A. (2001). *IST 2000 R: Intelligenz-Struktur-Test 2000 R*. Bern, Switzerland: Hogrefe, Verlag für Psychologie.

Arthur, W., & Day, D. V. (1994). Development of a short-form for the Raven

Bilker, W. B., Hansen, J. A., Brensinger, C. M., Richard, J., Gur, R. E., & Gur, R. C. (2012). Development of abbreviated nine-item forms of the Raven's standard progressive matrices test. *Assessment*, *19*(3), 354-369. 10.1177/1073191112446655

Dunn, L. M., & Dunn, L. M. (1965). Peabody picture vocabulary test--. https://doi.org/10.1037/t15145-000

Henmon, V. A. C. (1958). Manual for The Henmon-Nelson Tests of Mental Ability.

Jackson, D. N. (1984). Multidimensional aptitude battery manual. Port Huron, MI: Research Psychologists Press.

Jackson, D. N. (1998). *Multidimensional aptitude battery II: manual*. Port Huron, MI: Sigma Assessment Systems.

Krieshok, T. S., & Harrington, R. G. (1985). A Review of the Multidimensional Aptitude Battery. *Journal of Counseling & Development*, *64*(1), 87-89.

Measurement, 54, 394–403.

Petermann, F. (2014). Intelligenz-Struktur-Test. *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie*, *62*(3), 225-6.

Raven, J. (2003). Raven progressive matrices. In *Handbook of nonverbal assessment* (pp. 223-237). Boston, MA: Springer US. https://doi.org/10.1007/978-1-4615-0153-4_11

Raven, J. C., Raven, J., & Court, J. H. (1962). Advanced progressive matrices set II. Oxford Psychologists Press.

Review of *Henmon-Nelson Tests of Mental Ability* (1958). [Review of the software *Henmon-Nelson tests of mental ability,* by T. A. Lamke & M. J. Nelson]. *Journal of Consulting Psychology, 22*(3), 241. https://doi.org/10.1037/h0038444

Ryan, J. J., & Ward, L. C. (1999). Validity, reliability, and standard errors of measurement for two seven-subtest short forms of the Wechsler Adult Intelligence Scale—III. *Psychological Assessment*, *11*(2), 207. https://doi.org/10.1037/1040-3590.11.2.207

Thorndike, R. L. (1986). *The Stanford-Binet intelligence scale: Guide for administering and scoring*. Riverside Publishing Company.

Watson, C. G., Plemel, D., Schaefer, A., Raden, M., Alfano, A. M., Anderson, P. E., ... & Anderson, D. (1992). The comparative concurrent validities of the Shipley Institute of Living Scale and the Henmon-Nelson Tests of Mental ability. *Journal of clinical psychology*, *48*(2), 233-239. https://doi.org/10.1002/1097-4679(199203)48:2%3C233::AID-JCLP2270480215%3E3.0.CO;2-R

Wechsler, D. (1955). Wechsler adult intelligence scale--. *Archives of Clinical Neuropsychology*.

Wechsler, D. (1981). WAIS-R: Manual: Wechsler adult intelligence scale-revised.

Williams, J. (1999). Test Review: Peabody Picture Vocabulary Test-(PPVT-III). *Canadian Journal of School Psychology*, *14*(2), 67-73. https://doi.org/10.1177/082957359901400 2

Youngstrom, E. A., Glutting, J. J., Watkins, M. W., Reynolds, C. R., & Kamphaus, R. W. (2003). Stanford-Binet Intelligence Scale: (SB4): Evaluating the empirical bases for interpretations. *Handbook of psychological and educational assessment: Intelligence, aptitude, and achievement*, 217-242.

Appendix G

Description of publication bias detection methods and numeric results

Funnel plot/contour-enhanced funnel plot (Light & Pillemer, 1984). A funnel plot is a graphical representation of the relationship between the effect size of each study and its precision (usually measured by the standard error or sample size). The plot typically shows a symmetrical distribution of studies around the overall effect size estimate, with smaller studies scattered more widely due to their greater sampling variability. Asymmetry in the funnel plot can indicate the presence of publication bias, as smaller studies with non-significant or negative results may be missing from the plot. Additionally, contour- or power-enhanced funnel plots (Kossmeier et al., 2020), incorporate different colors to visually represent the statistical power. Warmer colors (e.g., red) indicate lower power, while colder colors (e.g., green) indicate larger study power.

Sterne and Egger's Regression Test (Sterne & Egger, 2005). Sterne and Egger's regression method uses a linear regression analysis, specifically regressing effect size estimates against its standard error (representing precision). The regression is weighted by the inverse variance of the effect estimate. In cases where the intercept significantly deviates from zero, the test signals potential asymmetry in the funnel plot. Asymmetric funnel plots may suggest the presence of a publication bias or small-study effects.

Trim-and-fill method (Duval & Tweedie, 2000). The trim-and-fill method involves a process where the study effects to the left and right of the summary effect size estimate are ranked based on their strength. Subsequently, these ranks are compared using a Wilcoxon-test, similar to the approach used for assessing funnel plot asymmetry (see Sterne and Egger's Regression test). In case of asymmetry, the smallest studies in the funnel plot are iteratively removed. Each removal is followed by a re-estimation of the summary effect ("trim-and-fill") (Borenstein et al., 2021). Asymmetry is then corrected by imputing the missing studies that are presumed to cause the asymmetry. Essentially, the method estimates the number of missing studies that would need to be added to make the funnel plot symmetric. These imputed studies are "filled" on the opposite side of the funnel plot to balance the plot. The adjusted summary effect size is then recalculated, incorporating both the original and the imputed studies. This iterative process continues until the funnel plot is symmetrical, providing a more accurate estimate of the true effect size by accounting for potential publication bias (Duval & Tweedie, 2000).

Rank Correlation method (Begg & Mazumdar 1994). The rank correlation method is a non-parametric correlation to detect asymmetries in the funnel plot. This method uses Kendall's τ to quantify the rank correlation between variables, specifically assessing the association between the standardized effect size and the variance (or standard error) of the study-specific effect estimates. A significant positive value of Kendall's τ indicates that smaller studies tend to report more extreme effect sizes, which is suggestive of publication bias. The Rank Correlation method is particularly useful because it does not assume a specific distribution of effect sizes, making it robust to various types of data. However, it is less powerful than some parametric tests and may not detect publication bias in the presence of high heterogeneity among studies (Begg & Mazumdar, 1994; Sterne & Egger, 2005).

Test of Excess of Significance (Ioannidis & Trikalinos, 2007). The Test of excess of significance is used to detect an overrepresentation of studies with statistically significant results. If publication bias is present, the observed effect sizes in the meta-analysis will be systematically larger than the expected effect sizes based on the individual study sample sizes and variances. To perform the test, the expected effect size for each study is calculated based on its sample size and variance, assuming no publication bias. The observed effect size is then compared to the expected effect size using a *z*-test. The *z*-score is calculated as the difference between the observed and expected effect sizes divided by the standard error of the expected effect size. If the *z*-score is greater than a certain threshold (usually 1.96, corresponding to a two-tailed *p*-value of 0.05), the study is considered to have an excess of significance, indicating that its observed effect size is larger than expected due to publication bias.

***p*-Curve (Simonsohn et al. 2014), *p*-uniform (van Assen et al., 2015), *p*-uniform* (van Aert & van Assen, 2021).** These three methods facilitate the estimation of the meta-analytical summary effects based solely on the distributions of *p*-value and associated degrees of freedom of statistically significant published primary studies. The rationale behind these methods is based on the idea that selective reporting is more influenced by statistical significance than the magnitude of the effect sizes. *P*-Curve serves as a diagnostic tool to examine evidence for both publication bias and *p*-hacking. It displays the distribution of statistically significant *p*-values ($p < .05$) within a study set, with the curve's shape depending on evidential value and sample size. A uniformly distributed *p*-curve suggests no evidential value under the null hypothesis, while a right-skewed curve indicates evidential value under the alternative hypothesis. Binomial tests and continuous tests (Stouffer method) are employed to assess right-skewness, while a 33% power test is used for left-skewed curves, indicating potential bias or *p*-hacking. Thresholds for bias indication consider combining full and half *p*-curve analyses. Both *p*-uniform and *p*-uniform* share similarities with *p*-curve but differ in their approaches to defining the fit to the uniform distribution. While *p*-curve and *p*-uniform rely on *p*-values of significant studies, *p*-uniform* considers both significant and non-significant studies. All three methods estimate an adjusted summary effect size and offer a means to calculate confidence intervals. The *p*-uniform method works through a process where the *p*-values of the individual studies are analyzed to detect and correct for publication bias. Specifically, the method identifies studies with statistically significant results and assumes that studies with non-significant results are underrepresented in the literature. By adjusting the distribution of *p*-values to fit a uniform distribution, *p*-uniform provides an unbiased estimate of the overall effect size (van Assen et al., 2015). *P*-uniform* is an enhanced version of *p*-uniform that also utilizes the distribution of *p*-values from the individual studies that are included in the meta-analysis. It adjusts for publication bias by estimating the effect size based on the assumption that the *p*-values of the studies should follow a uniform distribution under the null hypothesis. This adjustment allows for a more accurate estimation of the true effect size, even when there is heterogeneity among the studies. P-uniform* is particularly effective in addressing heterogeneity, making it a more precise tool for meta-analyses with varied study characteristics (van Aert & van Assen, 2021).

Specification Curve Analysis (SCA) (Simonsohn et al., 2015). The Specification Curve Analysis (SCA) provides a comprehensive approach to exploring the robustness of research findings by systematically examining the impact of different analytical choices on the results. SCA involves generating a "specification curve" by estimating the effect size for all possible combinations of reasonable analytical decisions, such as model specifications, variable selections, and data transformations. The process begins with defining a set of plausible analytical choices based on theoretical and empirical considerations. Each combination of these choices is then applied to the dataset to produce an array of results. The specification curve is a visual representation of these results, plotting the effect sizes against their corresponding specifications. This allows for the assessment of the consistency of their findings across a wide range of analytical decisions, thereby identifying which results are robust and which are sensitive to specific choices. By presenting a complete picture of how different specifications affect the results, SCA enhances the credibility and reproducibility of research findings. It also facilitates a better understanding of the robustness of empirical conclusions (Simonsohn et al., 2015).

Table S4. *Numeric outcomes of publication bias detection methods.*

Method	Full sample (k = *14*)	Trimmed sample (k = 13)
Trim-and-Fill	missing studies on the right side: 0 adjusted effect size = 0.131 p = .0005	missing studies on the right side: 3 adjusted effect size = 0.065 p = .027
Rank Correlation	τ = 0.429 p = .036	τ = 0.256 p = .252
Egger's regression	Z = -0.660 p = .522	Z = -0.201 p = .845
Test of Excess Significance	p = .999	p = .208
p-curve	continuous tests yielded p's < .0001 for evidential value indication and p = .719 for *p*-hacking indication; estimated statistical power: 99%	continuous tests yielded p's < .0001 for evidential value indication and p = .644 for *p*-hacking indication; estimated statistical power: 92%
p-uniform	estimate = 0.126 p = .003	estimate = 0.060 p = .009
p-uniform*	estimate = 0.139 p = .035	estimate = 0.081 p = .003

References Appendix G

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, 1088-1101. https://doi.org/10.2307/2533446

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2021). Introduction to meta-analysis. *John Wiley & Sons*.

Duval, S., & Tweedie, R. (2000). Trim and fill: a simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455-463. https://doi.org/10.1111/j.0006-341X.2000.00455.x

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *bmj*, *315*(7109), 629-634. https://doi.org/10.1136/bmj.315.7109.629

Ioannidis, J. P., & Trikalinos, T. A. (2007). An exploratory test for an excess of significant findings. *Clinical trials*, *4*(3), 245-253. https://doi.org/10.1177/1740774507079441

Kossmeier, M., Tran, U. S., & Voracek, M. (2020). Power-enhanced funnel plots for meta-analysis: The sunset funnel plot. *Zeitschrift fur Psychologie*, 228(1), 43. https://doi.org/10.1027/2151-2604/a000392

Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research*. Harvard University Press.

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Better P-curves: Making P-curve analysis more robust to errors, fraud, and ambitious P-hacking, a Reply to Ulrich and Miller (2015). *Journal of Experimental Psychology: General*, *144*(6), 1146–1152. https://doi.org/10.1037/xge0000104

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2020). Specification curve analysis. Nature Human Behaviour, 4(11), 1208-1214. https://doi.org/10.1038/s41562-020-0912-z

Sterne, J. A., & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. *Publication bias in meta-analysis: Prevention, assessment and adjustments*, 99-110. https://doi.org/10.1002/0470870168

van Aert, R. C., & van Assen, M. A. (2021). Correcting for publication bias in a meta-analysis with the *p*-uniform* method. https://doi.org/10.31222/osf.io/zqjr9

van Assen, M. A., van Aert, R., & Wicherts, J. M. (2015). Meta-analysis using effect size distributions of only statistically significant studies. *Psychological methods*, *20*(3), 293. https://doi.org/10.1037/met0000025

Appendix H

R-Code (full analysis; all studies included, k = 14)

```
# file needed: Coding.xlsx
################################################################################
####

# libraries ----
library(clubSandwich)
library(data.table)
library(plyr)
library(dplyr)
library(foreign)
library(fpc)
library(ggplot2)
library(ggpubr)
library(grid)
library(gridExtra)
library(gsl)
library(mclust)
library(haven)
library(Matrix)
library(metafor)
library(metaviz)
library(psychmeta)
library(puniform)
library(readxl)
library(robumeta)
library(SciViews)
library(stringr)
library(writexl)

# import data
data <- read_xlsx("Coding.xlsx")
data

# prepare data
data$sample <- as.factor(data$sample)
data$effect_id <- as.factor(data$effect_id)
data$design <- as.factor(data$design)
data$publication_status <- as.factor(data$publication_status)
data$publisher <- as.factor(data$publisher)
data$manuscript_type <- as.factor(data$manuscript_type)
data$peer_reviewed <- as.factor(data$peer_reviewed)
data$funding_yn <- as.factor(data$funding_yn)
data$previously_included <- as.factor(data$previously_included)
data$geographic_location <- as.factor(data$geographic_location)
data$sample_type <- as.factor(data$sample_type)
data$IQ_test <- as.factor(data$IQ_test)
data$attractiveness_type <- as.factor(data$attractiveness_type)
data$attractiveness_rating <- as.factor(data$attractiveness_rating)
```

```
data$raters <- as.factor(data$raters)
data$attractiveness_scale <- as.factor(data$attractiveness_scale)
data$age_mean <- as.numeric(data$age_mean)
data$percentage_male <- as.numeric(data$percentage_male)
data$IQ_mean <- as.numeric(data$IQ_mean)
data$IQ_SD <- as.numeric(data$IQ_SD)
data$IQ_reliability <- as.numeric(data$IQ_reliability)
data$p_value <- as.numeric(data$p_value)
data$r_se <- as.numeric(data$r_se)
data$r <- as.numeric(data$r)
data$V <- as.numeric(data$V)
data$N <- as.numeric(data$N)
data$z <- as.numeric(data$z)
data$z_se <- as.numeric(data$z_se)

# omit missing data
na.omit(data)
summary(data)

## effect size calculation ----
## calculate variance for ES and transform into standard error (Cooper et al., 2019)

# Variance of r. Computed as (1-r^2)^2/n-1
r <- data$r
n <- data$N

# Function to calculate variance of r
calculate_variance <- function(n, r) {
  (1 - r^2)^2 / (n - 1)
}

# Calculate variances for all studies
variances <- mapply(calculate_variance, n, r)

print(variances)

# Standard error of r. Computed as sqrt(V)
se <- data$variances

# Calculate standard errors
standard_errors <- sqrt(variances)
print(standard_errors)

# Print the standard errors
print(standard_errors)

# Standard error of z. Computed as 0.5*ln(1+r/1-r) or 1/sqrt(N-3)
r <- data$r
N <- data$N
z_se <- 1/sqrt(N-3)
print(z_se)
```

```
# Fisher's z transformation of r. Computed as atanh(r)
data_zcor <- escalc(measure = "ZCOR",
                    ri = r,
                    ni = N,
                    data = data)
data_zcor

########################
# Random Effects Model #
########################

# calculate Random Effects Model
res_data <- rma(yi, vi,
         data = data_zcor,
         method = "REML")
res_data

# convert z back to r
predict(res_data, digits = 3, transf = transf.ztor)

# calculate confidence interval for the amount of heterogeneity
confint(res_data)

######################
#### Forest Plot ####
######################

# Forest Plot
png(filename="Forest_Plot_full.png", res=350, width=3196, height=2448)
forest(res_data,
    slab = data_zcor$author_names,
    addcred = TRUE, #addcred: prediction interval
    col = "navyblue",
    xlab = "Fisher's z Transformed Correlation Coefficient",
    steps = 5)

# Add column headings to the plot
text(-3.8, 7.4, "Study name", pos = 4, cex = .9)
text(6.9, 3.2, "Pearson's r [95% CI]", pos = 3, cex = .9)

par(cex = .9, font = 1)
text(0.75, 15, "Pearson r [95% CI]", pos = 3, cex = 1)
dev.off()

######################
#### Baujat Plot ####
######################

# Baujat Plot (Baujat et al., 2002)
png(filename="Baujat_Plot_full.png", res=350, width=3196, height=2448)
baujat(res_data, symbol = "slab")
```

```
baujat
dev.off()

######################
##### GOSH Plot ######
######################

# GOSH-Plot (Olkin et al., 2012)
png(filename="Goshplot_full.png", res=350, width=2196, height=2048)
goshplot_full <- gosh(res_data)
plot(goshplot_full, cex = 0.3, out = 1, alpha = 0.9, col =
    c("#C8110a","#0063A6"))
dev.off()
print.gosh.rma(goshplot_full)

# Subgroup Forest Plot of Age Groups
png(filename="Rainforestplot_age_full.png", res=350, width=2196, height=2048)
viz_forest(x = res_data,
        study_labels = data_zcor$author_names,
        method = "REML",
        group = data$sample_type,
        summary_label = c("Adults", "Children", "Mixed"),
        xlab = "Correlation",
        variant = "rain",
        text_size = 2,
        annotate_CI = T)
viz_forest
dev.off()

# Subgroup Forest Plot of Attractiveness Assessment
png(filename="Rainforestplot_attr_full.png", res=350, width=2196, height=2048)
viz_forest(x = res_data,
      study_labels = data_zcor$author_names,
      method = "REML",
      group = data$attractiveness_type,
      summary_label = c("Live", "Photographs"),
      xlab = "Correlation",
      variant = "rain",
      text_size = 2,
      annotate_CI = T)
viz_forest
dev.off()

######################
## Publication Bias ##
######################

# Funnel Plot
png(filename="Funnel_Plot_full.png", res=350, width=2196, height=2048)
viz_funnel(x = res_data,
      method = "REML",
```

```
          contours = T,
          sig_contours = T,
          trim_and_fill = TRUE,
          trim_and_fill_side = "left",
          egger = TRUE,
          x_trans_function = tanh)
viz_funnel
dev.off()

# Power-enhanced Funnel Plot (Kossmeier et al., 2020)
png(filename="Powerenhanced_Plot_full.png", res=350, width=2196, height=2448)
viz_sunset(
  res_data,
  y_axis = "se",
  true_effect = NULL,
  method = "DL",
  sig_level = 0.05,
  power_stats = TRUE,
  power_contours = "discrete",
  contours = FALSE,
  sig_contours = TRUE,
  text_size = 3,
  point_size = 2,
  xlab = "Effect",
  ylab = NULL,
  x_trans_function = NULL,
  x_breaks = NULL,
  y_breaks = NULL,
  x_limit = NULL,
  y_limit = NULL
)
viz_sunset
dev.off()

# Power-enhanced Funnel Plot
png(filename="Contourenhanced_Plot_full.png", res=350, width=2196, height=2448)
funnel(res_data, level=c(90, 95, 99), refline=0, legend=TRUE)
funnel
dev.off()

# Egger's Regression Test (Sterne & Egger, 2005)
regtest(res_data,
       model = "lm",
       predictor = "sei")

# Rank Correlation Method (Begg & Mazumdar, 1994)
ranktest(res_data)

# p-uniform effect size estimation (van Assen et al., 2015)
puniform(ri = data_zcor$r,
        ni = data_zcor$N,
```

```
        alpha = .05,
        side = "right",
        method = "P",
        plot = TRUE)

# p-uniform* (van Aert & van Assen, 2021)
punistar_data <- puni_star(ri = data_zcor$r,
                      ni = data_zcor$N,
                      alpha = 0.1,
                      side = "right",
                      method = "ML")
punistar_data

# p-curve (Simonsohn et al., 2014) calculated with web app at p-curve.com

# Test of Excess of Significance
toe <- tes(data_zcor$r,
        data_zcor$V)
toe

# Trim-and-Fill-Method (Duval & Tweedie, 2000)
tf <- trimfill(res_data)
summary(tf)

########################
# Sensitivity Analyses #
########################

# Leave-one-out Sensitivity Analysis
leave1out.rma.uni(res_data, digits = 3)
leave1out

#########################
#### Meta-Regression ####
#########################

# publication year
pub_reg <- rma(yi, vi,
           mods = ~publication_year,
           data = data_zcor,
           method = "REML",
           digits = 3)
pub_reg
# b = -.012, p = .018, R^2 = 26.71%, tau^2 = .011 (SE = .006), tau = .105, I^2 = 96.41%, Q
(df = 12) = 885.785, p < .001

# mean age
age_reg <- rma(yi, vi,
           mods = ~age_mean,
           data = data_zcor,
           method = "REML",
```

```
            digits = 3)

###########################
### Moderator Analysis ###
###########################

# Sample type (children vs. adults vs. mixed)
data$sample_type <- factor(data$sample_type, levels =
                       c("adults", "children", "mixed"))

mod_age <- rma(yi, vi, mods = ~ sample_type, data = data_zcor,
                  method =
                 "REML")
summary(mod_age)

# Attractiveness assessment (live vs. photographs)
mod_attr <- rma(yi, vi, mods = ~ attractiveness_type,
                  data = data_zcor,
                  method =
                   "REML")
summary(mod_attr)

############################################
## Adults-only & Children-only samples ###
############################################

## Adult Samples (k = 8)
data_adults <- data[-c(1, 2, 4, 7, 8, 9), ]
summary(data_adults)

# Fisher's z transformation of r. Computed as atanh(r)
data_zcor_adults <- escalc(measure = "ZCOR",
                    ri = r,
                    ni = N,
                    data = data_adults)
data_zcor_adults

# calculate Random Effects Model for Adult Samples Only
res_data_adults <- rma(yi, vi,
                  data = data_zcor_adults,
                  method = "REML")
res_data_adults

# convert z back to r
predict(res_data_adults, digits = 3, transf = transf.ztor)

# calculate confidence interval for the amount of heterogeneity
confint(res_data_adults)

## Children Samples (k = 5)
data_children <- data[-c(2, 3, 5, 6, 10, 11, 12, 13, 14), ]
```

```
summary(data_children)

# Fisher's z transformation of r. Computed as atanh(r)
data_zcor_kids <- escalc(measure = "ZCOR",
                  ri = r,
                  ni = N,
                  data = data_children)
data_zcor_kids

# calculate Random Effects Model
res_data_kids <- rma(yi, vi,
               data = data_zcor_kids,
               method = "REML")
res_data_kids

# convert z back to r
predict(res_data_kids, digits = 3, transf = transf.ztor)

# calculate confidence interval for the amount of heterogeneity
confint(res_data_kids)

## Mixed Age Sample (k = 1)
data_mixed <- data[-c(1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14), ]
summary(data_mixed)

# Fisher's z transformation of r. Computed as atanh(r)
data_zcor_mixed <- escalc(measure = "ZCOR",
                  ri = r,
                  ni = N,
                  data = data_mixed)
data_zcor_mixed

# calculate Random Effects Model
res_data_mixed <- rma(yi, vi,
               data = data_zcor_mixed,
               method = "REML")
res_data_mixed

# convert z back to r
predict(res_data_mixed, digits = 3, transf = transf.ztor)

# calculate confidence interval for the amount of heterogeneity
confint(res_data_mixed)

## Live assessment Samples (k = 3)
data_live <- data[-c(4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14), ]
summary(data_live)

# Fisher's z transformation of r. Computed as atanh(r)
data_zcor_live <- escalc(measure = "ZCOR",
                  ri = r,
```

```
                    ni = N,
                    data = data_live)
data_zcor_live

# calculate Random Effects Model for Live Samples Only
res_data_live <- rma(yi, vi,
                 data = data_zcor_live,
                 method = "REML")
res_data_live

# convert z back to r
predict(res_data_live, digits = 3, transf = transf.ztor)

# calculate confidence interval for the amount of heterogeneity
confint(res_data_live)

## Photograph assessment Samples (k = 11)
data_photo <- data[-c(1, 2, 3), ]
summary(data_photo)

# Fisher's z transformation of r. Computed as atanh(r)
data_zcor_photo <- escalc(measure = "ZCOR",
                  ri = r,
                  ni = N,
                  data = data_photo)
data_zcor_photo

# calculate Random Effects Model for Photograph Samples Only
res_data_photo <- rma(yi, vi,
                data = data_zcor_photo,
                method = "REML")
res_data_photo

# convert z back to r
predict(res_data_photo, digits = 3, transf = transf.ztor)

# calculate confidence interval for the amount of heterogeneity
confint(res_data_photo)
```

Specification Curve Analysis

```
# file needed: Coding.xlsx
#
# Note: This script follows the code supplemented to Voracek, Kossmeier & Tran (2019)
# "Which Data to Meta-Analyze and How?"; https://doi.org/10.1027/2151-2604/a000357
# Code available at https://osf.io/nkv46/
################################################################################
####
```

```
#####################################
###### Specification Curve Analysis ######
#####################################

# libraries ----
library(readxl)
library(metafor)
library(robumeta)
library(clubSandwich)
library(stringr)
library(writexl)

# load meta-analytic data
x <- read_xlsx("Coding.xlsx")

# prepare data
x$sample_type <- as.factor(x$sample_type)
x$attractiveness_type <- as.factor(x$attractiveness_type)

# Construct grouping factors (internal, which factors)
sample_type <- c("children", "adults", "mixed")
attractiveness_type <- c("live", "photographs")

# different methods (external, how factors)
effect <- c("r", "z")
ma_method <- c("FE","REML", "unweighted")

# Construct all possible combinations of internal and external factors
specifications <- expand.grid(sample_type = sample_type,
                              attractiveness_type = attractiveness_type,
                              effect = effect,
                              ma_method = ma_method)
specifications <- data.frame(specifications,
                             mean = rep(NA, nrow(specifications)),
                             lb = rep(NA, nrow(specifications)),
                             ub = rep(NA, nrow(specifications)),
                             p = rep(NA, nrow(specifications)),
                             k = rep(NA, nrow(specifications)))

# Conduct specification analyses
for(i in 1:nrow(specifications)) {
  dat <- x
  ##########################################################################
  #### Determine specificaiton subset by using "Which" factors #######
  ##########################################################################
  if(specifications$sample_type[i] == "children") {
    dat <- dat[dat$sample_type == "children", ]
  } else {
    if(specifications$sample_type[i] == "adults") {
      dat <- dat[dat$sample_type == "adults", ]
    }
```

```
}
  if(specifications$attractiveness_type[i] == "live") {
    dat <- dat[dat$attractiveness_type == "live", ]
  } else {
 if(specifications$attractiveness_type[i] == "photographs") {
  dat <- dat[dat$attractiveness_type == "photographs", ]
  }
 }
 # only compute meta-analytic summary effects for specification subsets with at least two
studies/samples.
 if(nrow(dat) < 2) next

 # Save which study/sample IDs were selected by the "Which" factors for a given
specification.
 specifications$set[i] <- paste(rownames(dat), collapse = ",")
 ##########################################################################
 #### Determine specification analyses by using "How" factors ##########
 ##########################################################################
 if(specifications$effect[i] == "r") {
  #################### effect = r ################
  if(specifications$ma_method[i] == "FE") {
    mod <- rma(yi = dat$r, sei = dat$r_se, method = "FE")
  } else {
    if(specifications$ma_method[i] == "REML") {
      mod <- rma(yi = dat$r, sei = dat$r_se, method = "REML",
             control = list(stepadj=0.5, maxiter = 2000))
     } else {
       if(specifications$ma_method[i] == "unweighted") {
        mod <- rma(yi = dat$r, sei = dat$r_se, method = "FE",
                weights = 1/nrow(dat))
        }
     }
  }
  specifications$mean[i] <- mod$b[[1]]
  specifications$lb[i] <- mod$ci.lb[[1]]
  specifications$ub[i] <- mod$ci.ub[[1]]
  specifications$p[i] <- mod$pval[[1]]
  specifications$k[i] <- nrow(dat)
  ################################################
 } else {
  if(specifications$effect[i] == "z") {
   #################### effect = z ################
   if(specifications$ma_method[i] == "FE") {
    mod <- rma(yi = dat$z, sei = dat$z_se, method = "FE")
   } else {
    if(specifications$ma_method[i] == "REML") {
    mod <- rma(yi = dat$z, sei = dat$z_se, method = "REML",
            control = list(stepadj=0.5, maxiter = 2000))
     } else {
       if(specifications$ma_method[i] == "unweighted") {
        mod <- rma(yi = dat$z, sei = dat$z_se, method = "FE",
```

```
                    weights = 1/nrow(dat))
          }
        }
      }
      specifications$mean[i] <- tanh(mod$b[[1]]) # inverse of z
      specifications$lb[i] <- tanh(mod$ci.lb[[1]])
      specifications$ub[i] <- tanh(mod$ci.ub[[1]])
      specifications$p[i] <- mod$pval[[1]]
      specifications$k[i] <- nrow(dat)
      ################################################

# Only keep specifications with at least 2 studies/samples
specifications_full <- specifications[complete.cases(specifications),]
# Only keep unique study/sample subsets resulting from "Which" factor combinations.
specifications_full <- specifications_full[!duplicated(specifications_full[,
                         c("mean", "set", "ma_method", "effect")]), ]
# Indicator if all studies are included in the set
specifications_full$full_set <- as.numeric(specifications_full$set ==
                          paste(1:nrow(x), collapse =",", sep = ""))
    }
  }
}
# Save specification data
write.csv2(file = "specifications_IQAttr_complete.csv", specifications_full)
```

Specification Curve Analysis Plots

```
# file needed: specifications_IQAttr_complete.csv
#
# Note: This script follows the code supplemented to Voracek, Kossmeier & Tran (2019)
# "Which Data to Meta-Analyze and How?"; https://doi.org/10.1027/2151-2604/a000357
# Code available at https://osf.io/nkv46/
###############################################################################
####

####################################
## Descriptive Specification Plot ##
####################################

# libraries ----
library(readxl)
library(metafor)
library(ggplot2)
library(plyr)
library(ggpubr)
library(grid)
library(gridExtra)

# load data
specifications_full <- read.csv2('specifications_IQAttr_complete.csv')
```

```
specifications_full

# prepare data
specifications_full$mean <- as.numeric(specifications_full$mean)
specifications_full$lb <- as.numeric(specifications_full$lb)
specifications_full$ub <- as.numeric(specifications_full$ub)
specifications_full$p <- as.numeric(specifications_full$p)
specifications_full$k <- as.numeric(specifications_full$k)
specifications_full$set <- as.factor(specifications_full$set)

# prepare plotting data for tile plot
# grouping factors
sample_type <- c("children", "adults", "mixed")
attractiveness_type <- c("live", "photographs")

# different methods
effect <- c("r", "z")
ma_method <- c("FE", "REML", "unweighted")

x_rank <- rank(specifications_full$mean, ties.method = "random")
yvar <- rep(factor(rev(c(sample_type, attractiveness_type, effect, ma_method)),
                   levels = rev(c(sample_type, attractiveness_type,
                                  effect, ma_method))),
            times = nrow(specifications_full))
xvar <- rep(x_rank, each = length(levels(yvar)))
spec <- NULL

# Determine which specifications are observed and which are not
for(i in 1:nrow(specifications_full)) {
  id <- as.numeric(levels(yvar) %in% as.character
                   (unlist(specifications_full[i, 1:9])))
  spec <- c(spec, id)
}

plotdata <- data.frame(xvar, yvar, spec)

ylabels <- rev(c("sample: children", "sample: adults", "sample: mixed",
                 "live judgment", "photograph judgment",
                 "metric: r",
                 "metric: z",
                 "model: FE",
                 "model: REML",
                 "model: unweighted"
))

plotdata$k <- rep(specifications_full$k, each = length(levels(yvar)))
plotdata$fill <- as.factor(plotdata$k * plotdata$spec)
cols <- RColorBrewer::brewer.pal(min(11, length(levels(plotdata$fill)) - 1),
                                 "Spectral")
```

```
# Create specification tile plot ----------------------------------------
p_spec <- ggplot(data = plotdata, aes(x = xvar, y = as.factor(yvar),
                                      fill = fill)) +
  geom_raster() +
  geom_hline(yintercept = c(5, 7) + 0.5) +
  scale_x_continuous(position = "bottom") +
  scale_y_discrete(labels = ylabels) +
  scale_fill_manual(values = c("white",
                c(cols[floor(seq(from = 1, to = length(cols), length.out =
                                 length(levels(plotdata$fill)) - 1))]))) +
  labs(x = "Specification number", y = "Which/How factors") +
coord_cartesian(expand = F, xlim = c(0.5, nrow(specifications_full) + 0.5)) +
  theme_bw() +
  theme(legend.position = "none",
        axis.text = element_text(colour = "black"),
        axis.ticks = element_line(colour = "black"),
        plot.margin = margin(t = 5.5, r = 5.5, b = 5.5, l = 5.5, unit = "pt"))

# Create summary forest plot ----------------------------------------------------
specifications_full$xvar <- x_rank
yrng <- range(c(0, specifications_full$lb, specifications_full$ub))
ylimit <- c(yrng[1] - diff(yrng)*0.1, yrng[2] + diff(yrng)*0.1)

y_breaks_forest <- round(seq(from = round(ylimit[1], 1), to = round(ylimit[2],
                                                    1), by = 0.1), 2)
y_labels_forest <- format(y_breaks_forest, nsmall = 2)
y_breaks_forest <- c(ylimit[1], y_breaks_forest)
y_labels_forest <- c(ylabels[which.max(nchar(ylabels))], y_labels_forest)

p_forest <-
  ggplot(data = specifications_full, aes(x = xvar, y = mean)) +
  geom_errorbar(aes(ymin = lb, ymax = ub, col = as.factor(k)), width = 0,
              size = 0.25) +
  geom_line(col = "black", size = 0.25) +
  geom_hline(yintercept = 0, linetype = 2, size = 0.25) +
  scale_x_continuous(name = "") +
  scale_y_continuous(name = "Summary effect (r)", breaks = y_breaks_forest,
                 labels = y_labels_forest) +
  scale_color_manual(values = c(cols[floor(seq(from = 1, to = length(cols),
        length.out = length(levels(as.factor(specifications_full$k)))))])) +

  coord_cartesian(ylim = ylimit, xlim = c(0.5, nrow(specifications_full)
                                 + 0.5),
                expand = FALSE) +
  ggtitle("Intelligence and Attractiveness (k = 14)") +
  theme_bw() +
  theme(legend.position = "none",
        axis.text.x = element_blank(),
        axis.ticks.x = element_blank(),
        axis.text.y = element_text(colour = c("white", rep("black",
                                times = length(y_labels_forest) - 1))),
```

```
        axis.ticks.y = element_line(colour = c("white", rep("black",
                                    times = length(y_breaks_forest) - 1))),
        panel.grid.major.x = element_blank(),
        panel.grid.minor.x = element_blank(),
        panel.grid.major.y = element_line(),
        panel.grid.minor.y = element_blank(),
        plot.margin = margin(t = 5.5, r = 5.5, b = -15, l = 5.5, unit = "pt"))

# Create subset size indicator -------------------------------------------
yrng <- range(c(2, max(specifications_full$k)))
ylimit <- c(2, max(specifications_full$k))

y_breaks_size <- round(seq(from = yrng[1], to = yrng[2], by = 100), 0)
y_labels_size <- format(y_breaks_size, nsmall = 0)
y_breaks_size <- c(ylimit[1], y_breaks_size)
y_labels_size <- c(ylabels[which.max(nchar(ylabels))], y_labels_size)

p_size <-
  ggplot(data = specifications_full, aes(x = xvar, y = k)) +
  geom_area(fill = "gray35", color = "black", size = 0.25) +
  scale_x_continuous(name = "") +
  scale_y_continuous(name = "# Samples", breaks = y_breaks_size,
                     labels = y_labels_size) +
  coord_cartesian(ylim = ylimit, xlim = c(0.5, nrow(specifications_full)
                                     + 0.5),
                  expand = FALSE) +
  theme_bw() +
  theme(legend.position = "none",
        axis.text.x = element_blank(),
        axis.ticks.x = element_blank(),
        axis.text.y = element_text(colour = c("white", rep("black",
                                     times = length(y_labels_size) - 1))),
        axis.ticks.y = element_line(colour = c("white", rep("black",
                                    times = length(y_breaks_size) - 1))),
        panel.grid.major.x = element_blank(),
        panel.grid.minor.x = element_blank(),
        panel.grid.major.y = element_line(),
        panel.grid.minor.y = element_blank(),
        plot.margin = margin(t = 5.5, r = 5.5, b = -15, l = 5.5, unit = "pt"))

# Combine specfication tile plot, subset size indicator and forest plot
p <- gridExtra::arrangeGrob(p_spec, p_size, p_forest,
                            layout_matrix = matrix(c(3, 3, 3, 2, 1, 1, 1, 1, 1),
                                                   ncol = 1))
p <- ggpubr::as_ggplot(p)
ggsave("plot_SCA_IQAttr_complete.tiff", p, width = 16*1.2, height = 12*1.2,
       dpi = 1200, units = "cm", compression = "lzw+p")
```

R-Code

(trimmed analysis; Kanazawa (2011) – Group A excluded, k = 13)

```
# file needed: Coding.xlsx
################################################################################
####

# libraries ----
library(clubSandwich)
library(data.table)
library(plyr)
library(dplyr)
library(foreign)
library(fpc)
library(ggplot2)
library(ggpubr)
library(grid)
library(gridExtra)
library(gsl)
library(mclust)
library(haven)
library(Matrix)
library(metafor)
library(metaviz)
library(psychmeta)
library(puniform)
library(readxl)
library(robumeta)
library(SciViews)
library(stringr)
library(writexl)

# import data
data_k <- read_xlsx("Coding.xlsx")
data_k

# Remove Sample with exceptionally high correlation r = .381, Kanazawa (2011) - Group A
data_k <- data_k[-1, ]
summary(data_k)

# prepare data
data_k$sample <- as.factor(data_k$sample)
data_k$effect_id <- as.factor(data_k$effect_id)
data_k$design <- as.factor(data_k$design)
data_k$publication_status <- as.factor(data_k$publication_status)
data_k$publisher <- as.factor(data_k$publisher)
data_k$manuscript_type <- as.factor(data_k$manuscript_type)
data_k$peer_reviewed <- as.factor(data_k$peer_reviewed)
data_k$funding_yn <- as.factor(data_k$funding_yn)
data_k$previously_included <- as.factor(data_k$previously_included)
data_k$geographic_location <- as.factor(data_k$geographic_location)
```

```
data_k$sample_type <- as.factor(data_k$sample_type)
data_k$IQ_test <- as.factor(data_k$IQ_test)
data_k$attractiveness_type <- as.factor(data_k$attractiveness_type)
data_k$attractiveness_rating <- as.factor(data_k$attractiveness_rating)
data_k$raters <- as.factor(data_k$raters)
data_k$attractiveness_scale <- as.factor(data_k$attractiveness_scale)
data_k$age_mean <- as.numeric(data_k$age_mean)
data_k$percentage_male <- as.numeric(data_k$percentage_male)
data_k$IQ_mean <- as.numeric(data_k$IQ_mean)
data_k$IQ_SD <- as.numeric(data_k$IQ_SD)
data_k$IQ_reliability <- as.numeric(data_k$IQ_reliability)
data_k$p_value <- as.numeric(data_k$p_value)
data_k$r_se <- as.numeric(data_k$r_se)
data_k$r <- as.numeric(data_k$r)
data_k$V <- as.numeric(data_k$V)
data_k$N <- as.numeric(data_k$N)
data_k$z <- as.numeric(data_k$z)
data_k$z_se <- as.numeric(data_k$z_se)

# omit missing data_k
na.omit(data_k)
summary(data_k)

# Fisher's z transformation of r. Computed as atanh(r)
data_zcor_k <- escalc(measure = "ZCOR",
                ri = r,
                ni = N,
                data = data_k)
data_zcor_k

########################
# Random Effects Model #
########################

# calculate Random Effects Model
res_data_k <- rma(yi, vi,
             data = data_zcor_k,
             method = "REML")
res_data_k

# convert z back to r
predict(res_data_k, digits = 3, transf = transf.ztor)

# calculate confidence interval for the amount of heterogeneity
confint(res_data_k)

######################
#### Forest Plot ####
######################

# Forest Plot
```

```
png(filename="Forest_Plot_trimmed.png", res=350, width=3196, height=2448)
forest(res_data_k,
    slab = data_zcor_k$author_names,
    addcred = TRUE, #addcred: prediction interval
    col = "navyblue",
    xlab = "Fisher's z Transformed Correlation Coefficient",
    steps = 5)

# Add column headings to the plot
text(-3.8, 7.4, "Study name", pos = 4, cex = .9)
text(6.9, 3.2, "Pearson's r [95% CI]", pos = 2, cex = .9)

par(cex = .9, font = 1)
text(0.75, 15, "Pearson r [95% CI]", pos = 1, cex = 1)
dev.off()

######################
#### Baujat Plot ####
######################

# Baujat Plot (Baujat et al., 2002)
png(filename="Baujat_Plot_trimmed.png", res=350, width=3196, height=2448)
baujat(res_data_k, symbol = "slab")
baujat
dev.off()

######################
##### GOSH Plot ######
######################

# GOSH-Plot (Olkin et al., 2012)
png(filename="Goshplot_trimmed.png", res=350, width=2196, height=2048)
goshplot_trimmed <- gosh(res_data_k)
plot(goshplot_trimmed, cex = 0.3, out = 1, alpha = 0.9, col =
    c("#C8110a","#0063A6"))
dev.off()
print.gosh.rma(goshplot_trimmed)

# Subgroup Forest Plot of Age Groups
png(filename="Rainforestplot_age_trimmed.png", res=350, width=2196, height=2048)
viz_forest(x = res_data,
      study_labels = data_zcor_k$author_names,
      method = "REML",
      group = data$sample_type,
      summary_label = c("Adults", "Children", "Mixed"),
      xlab = "Correlation",
      variant = "rain",
      text_size = 2,
      annotate_CI = T)
viz_forest
dev.off()
```

```
# Subgroup Forest Plot of Attractiveness Assessment
png(filename="Rainforestplot_attr_trimmed.png", res=350, width=2196,
   height=2048)
viz_forest(x = res_data,
        study_labels = data_zcor_k$author_names,
        method = "REML",
        group = data$attractiveness_type,
        summary_label = c("Live", "Photographs"),
        xlab = "Correlation",
        variant = "rain",
        text_size = 2,
        annotate_CI = T)
viz_forest
dev.off()

######################
## Publication Bias ##
######################

# Funnel Plot
png(filename="Funnel_Plot_trimmed.png", res=350, width=2196, height=2048)
viz_funnel(x = res_data_k,
        method = "REML",
        contours = T,
        sig_contours = T,
        trim_and_fill = TRUE,
        trim_and_fill_side = "left",
        egger = TRUE,
        x_trans_function = tanh)
viz_funnel
dev.off()

# Power-enhanced Funnel Plot (Kossmeier et al., 2020)
png(filename="Powerenhanced_Plot_trimmed.png", res=350, width=2196,
   height=2448)
viz_sunset(
  res_data_k,
  y_axis = "se",
  true_effect = NULL,
  method = "DL",
  sig_level = 0.05,
  power_stats = TRUE,
  power_contours = "discrete",
  contours = FALSE,
  sig_contours = TRUE,
  text_size = 3,
  point_size = 2,
  xlab = "Effect",
  ylab = NULL,
  x_trans_function = NULL,
  x_breaks = NULL,
```

```
  y_breaks = NULL,
  x_limit = NULL,
  y_limit = NULL
)
viz_sunset
dev.off()

# Power-enhanced Funnel Plot
png(filename="Contourenhanced_Plot_trimmed.png", res=350, width=2196,
   height=2448)
funnel(res_data_k, level=c(90, 95, 99), refline=0, legend=TRUE)
funnel
dev.off()

# Egger's Regression Test (Sterne & Egger, 2005)
regtest(res_data_k,
        model = "lm",
        predictor = "sei")

# Rank Correlation Method (Begg & Mazumdar, 1994)
ranktest(res_data_k)

# p-uniform effect size estimation (van Assen et al., 2015)
puniform(ri = data_zcor_k$r,
         ni = data_zcor_k$N,
         alpha = .05,
         side = "right",
         method = "P",
         plot = TRUE)

# p-uniform* (van Aert & van Assen, 2021)
punistar_data <- puni_star(ri = data_zcor_k$r,
                           ni = data_zcor_k$N,
                           alpha = 0.1,
                           side = "right",
                           method = "ML")
punistar_data

# Test of Excess of Significance
toe_k <- tes(data_zcor_k$r,
           data_zcor_k$V)
toe_k

# Trim-and-Fill-Method (Duval & Tweedie, 2000)
tf <- trimfill(res_data_k)
summary(tf)

########################
# Sensitivity Analyses #
########################
```

```
# Leave-one-out Sensitivity Analysis
data_leave1out <- leave1out(res_data_k,
                            digits = 3)
data_leave1out

##########################
#### Meta-Regression ####
##########################

# publication year
pub_reg_k <- rma(yi, vi,
                 mods = ~publication_year,
                 data = data_zcor_k,
                 method = "REML",
                 digits = 3)
pub_reg_k

# mean age
age_reg_k <- rma(yi, vi,
                 mods = ~age_mean,
                 data = data_zcor_k,
                 method = "REML",
                 digits = 3)
age_reg_k

###########################
### Moderator Analysis ###
###########################

# sample type (adults vs children vs mixed)
data_k$sample_type <- factor(data_k$sample_type, levels =
                             c("adults", "children", "mixed"))

mod_age_k <- rma(yi, vi, mods = ~ sample_type, data = data_zcor_k,
                 method =
                  "REML")
summary(mod_age_k)

# attractiveness assessment (live vs photographs)
mod_attr_k <- rma(yi, vi, mods = ~ attractiveness_type,
                  data = data_zcor_k,
                  method =
                   "REML")
summary(mod_attr_k)
```

Appendix I

Documentation of literature search

Overview

A total of four databases were searched for relevant articles: PubMed, Scopus, ISI Web of Science, Open Access Thesis and Dissertations. A cited reference search was performed for Kanazawa & Kovar (2004) via ISI Web of Science. Studies were screened (titles and abstracts) from 05/08/2023 to 09/05/2023. Last checked on: 10/05/2024.

Classification of searched literature before coding full texts:

Documentation

1. **PubMed**
   Date (first search): 05/08/2023
   Hits: 1382
   Date (second search): 09/05/2023
   Hits: 1507
2. **Scopus**
   Date (first search): 05/08/2023
   Hits: 8598
   Hits (limited to psychology): 441
   Date (second search): 09/05/2023
   Hits (limited to psychology): 932
3. **ISI Web of Science**
   Date: 05/08/2023
   Hits: 2024
   Cited reference search: Kanazawa, S*
   Hits: 114
4. **Open Access Thesis and Dissertations**
   Date: 08/05/2023
   Hits: 748
   Hits (limited to Psychology): 14
5. **Backward Search**
   Additionally screened: 2

Appendix J

Additional results of psychometric meta-analysis

Figure 1A. Baujat-Plot of full dataset (*k* = 14). *Contribution of each study to the overall Q-test statistic for heterogeneity on the horizontal axis versus the influence of each study on the vertical axis.*

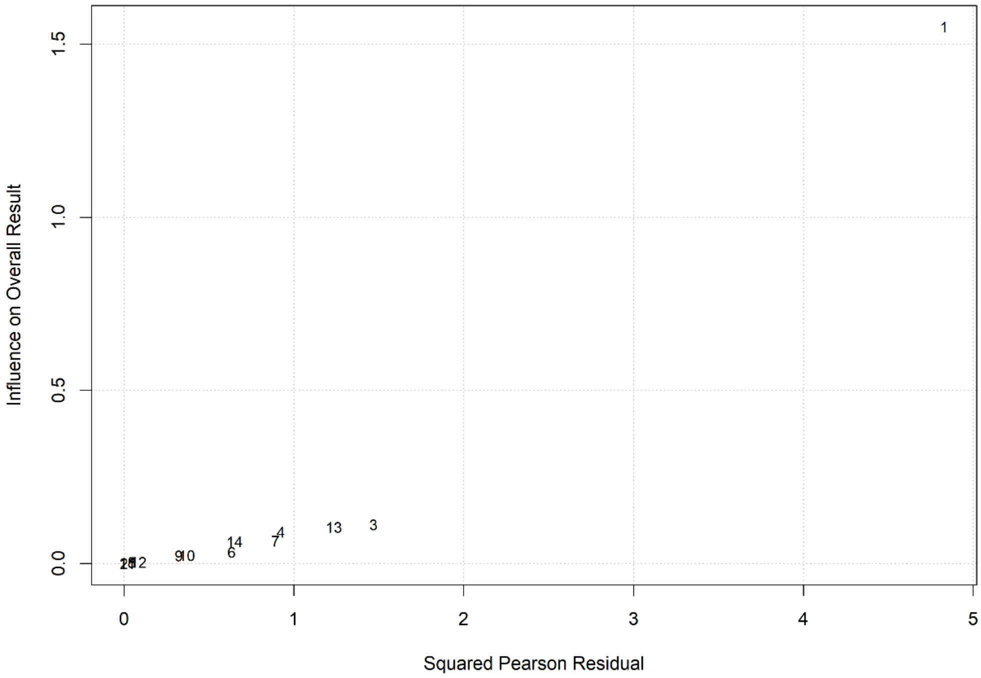


Figure 1B. Baujat-Plot of trimmed dataset (*k* = 13). *Contribution of each study to the overall Q-test statistic for heterogeneity on the horizontal axis versus the influence of each study on the vertical axis*

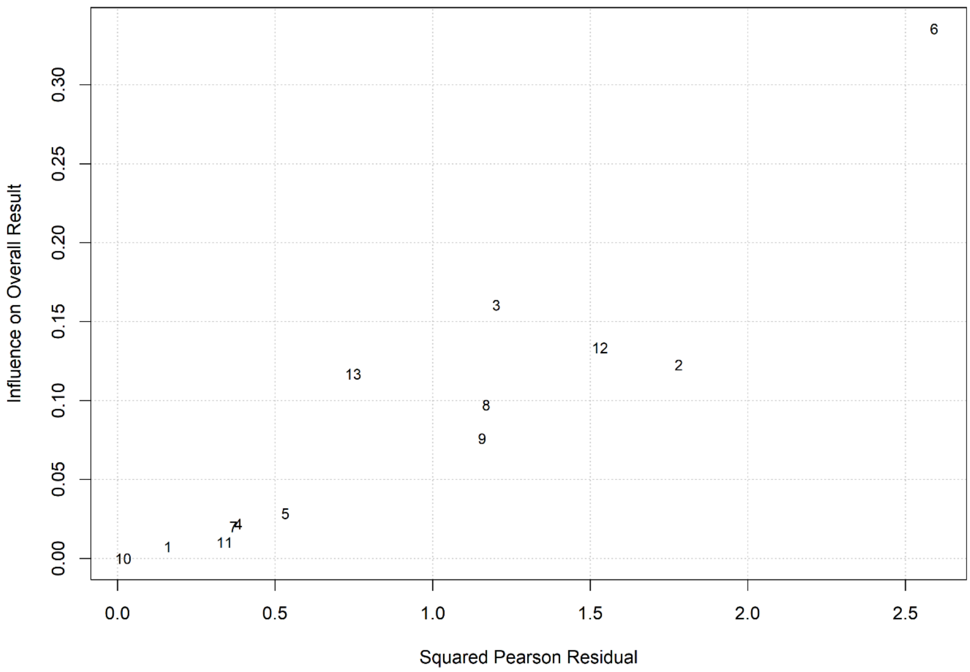


Figure 2A. Rainforest Plot for age groups (children vs. adults vs. mixed).

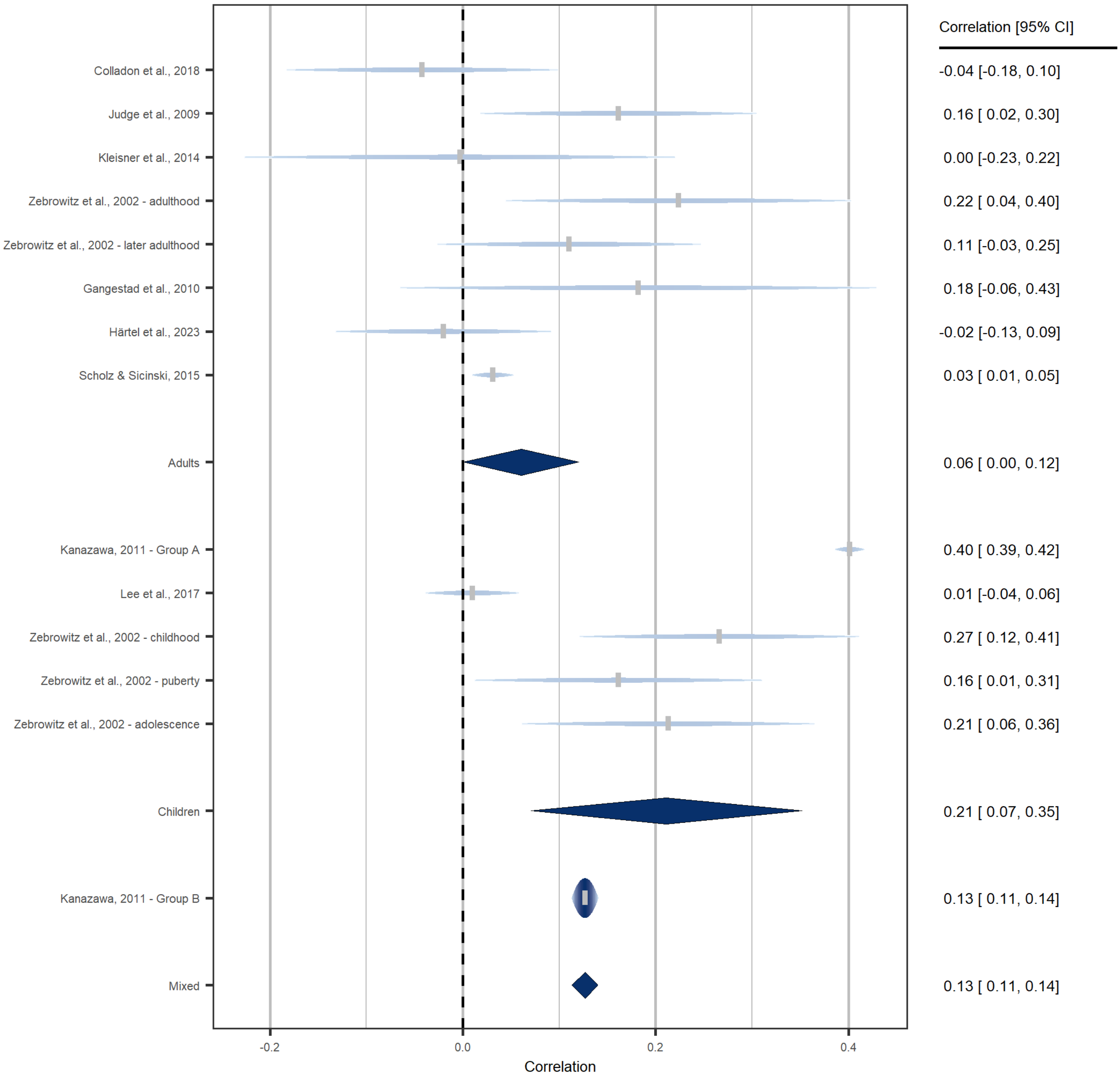
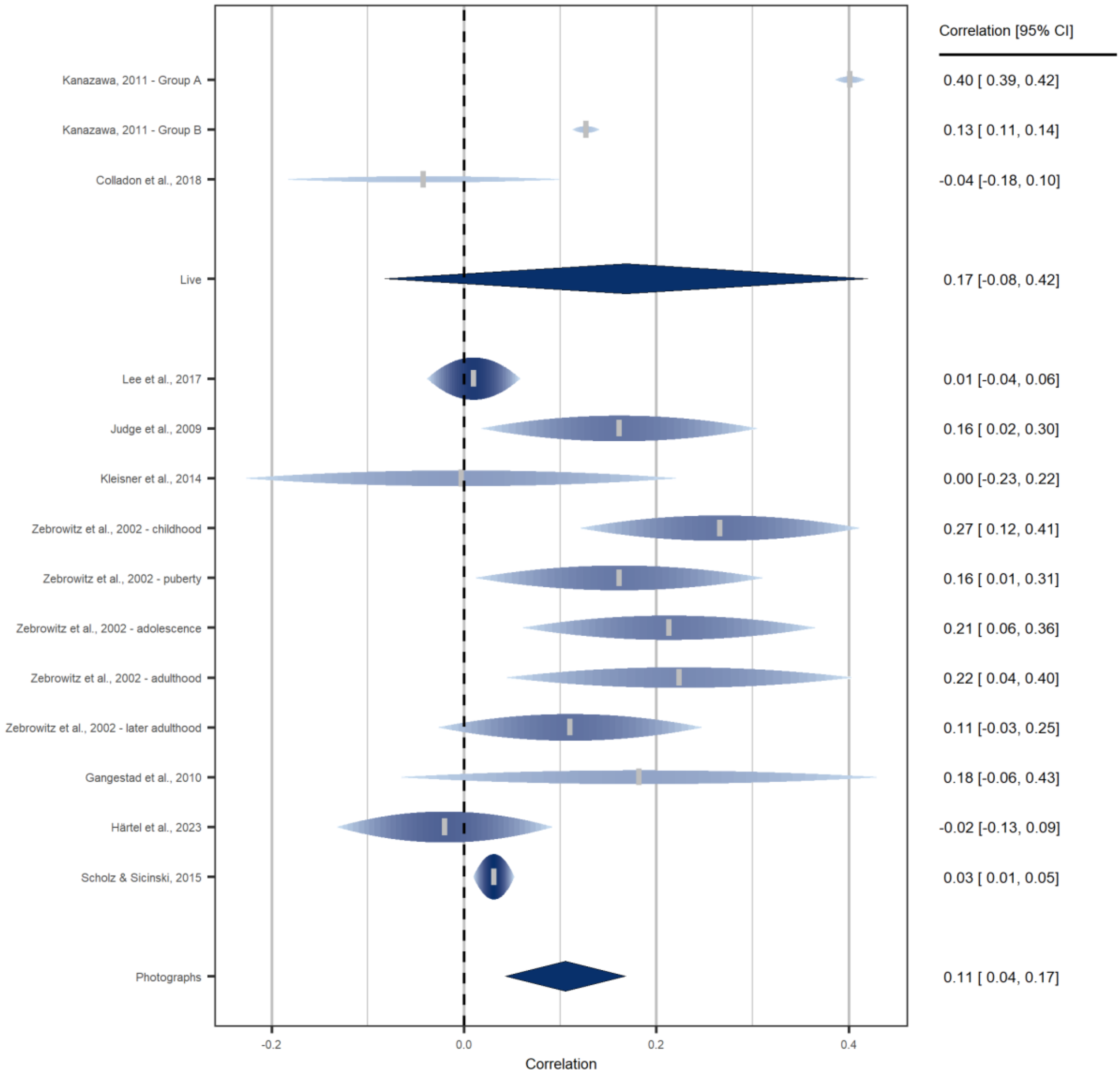


Figure 2B. Rainforest Plot for attractiveness assessment (live vs. photographs).



Abbreviations

Abbreviation	Meaning
CHC	Cattell-Horn-Carroll
ICHH	Immunocompetence handicap hypothesis
xAM	Cross-trait assortative mating
a/h^2	Genetic heritability factor
c	Common/shared environmental factor
e	Non-shared environmental factor
ACE model	Genetics (A), shared environmental influences (C), non-shared environmental influences (E)
IQ	Intelligence Quotient
ERP	Event-related potential
BFI	Big-Five Inventory
GWAS	Genome-Wide Association Study
SAT	Scholastic Assessment Test
GPA	Grade Point Average
MCFC	Males-compete/females-choose
MMC	Mutual mate choice
SCA	Specification Curve Analysis
GOSH	Graphical Display of Study Heterogeneity
WAIS	Wechsler Adult Intelligence Scale
RPM	Raven Progressive Matrices Test
MAB	Multidimensional Aptitude Battery
IST-R 2000	Intelligence Structure Test 2000-R
PPVT	Peabody Picture Vocabulary Test