OR-01 Mixed-sex teams and collective intelligence

The question of how the sex composition of a group affects its intelligence, creativity, and performance is among the most consequential and heavily debated topics in the organizational and behavioral sciences. As modern societies increasingly rely on team-based knowledge work, understanding the dynamics of mixed-sex groups is critical for optimizing human cooperation in couples, families, small groups, organizations, and societies at large. However, the empirical landscape is highly fractured. Experimental data from behavioral economics, psychometrics, and organizational psychology frequently yield contradictory results, heavily moderated by task demands, group sizes, and communication mediums.

To navigate this complexity, this report critically examines the exhaustive literature on mixed-sex teams and collective intelligence. It evaluates the foundational evidence for a "collective intelligence" factor, assesses the empirical record on gender composition and team performance, and delineates the specific conditions under which mixed-sex teams outperform or underperform single-sex groups.

The Three Layers of Human Interaction

To analyze team dynamics rigorously without falling into biological determinism or radical social constructivism, it is necessary to separate human interaction into three distinct, interacting layers: the physiological, the behavioral, and the cultural.

Layer 1: Physiology (Measurable Foundations)

At the physiological layer, researchers examine the cognitive and neurological hardware that facilitates social interaction and group coordination. The most relevant construct in the study of collective intelligence is "Theory of Mind" (ToM)—the cognitive capacity to attribute mental states to oneself and others, and to understand that others possess beliefs, desires, and intentions distinct from one's own.

The standard psychometric tool for assessing a specific, low-level component of ToM is the "Reading the Mind in the Eyes Test" (RMET), originally developed by Simon Baron-Cohen and colleagues. The RMET requires participants to identify complex emotional states solely from photographs of the eye region of faces. Performance on the RMET is frequently used as a proxy for "social sensitivity" or cognitive empathy.

The physiological and neurological ability to decode these visual cues exhibits a well-documented sex difference. In a comprehensive meta-analysis published in the *North American Journal of Psychology*, Kirkland et al. (2013) analyzed 42 studies comprising 4,290 healthy adults[1]. They found a statistically significant female advantage on the RMET, yielding a small-to-medium effect size (Hedges' $g = 0.177$, Cohen's $d = 0.47$)[1]. The authors noted that this female superiority was homogenous across studies, indicating a robust, cross-cultural baseline difference in this specific facet of social cognition[1].

This finding was subsequently replicated on an unprecedented scale, confirming its robustness. Greenberg et al. (2022), publishing in the *Proceedings of the National Academy of Sciences*

(PNAS), leveraged four massive datasets totaling 312,739 individuals across 57 countries. They confirmed a ubiquitous, on-average female advantage on the RMET across the lifespan and across almost all studied linguistic and cultural boundaries[2]. From a physiological and psychometric standpoint, the evidence is highly robust that women, on average, possess a higher baseline capacity for decoding nonverbal emotional cues from the face.

Layer 2: Behaviour (Statistical, Overlapping Distributions)

Physiological baselines influence, but do not strictly dictate, higher-order behavior. Behavioral economics and social psychology demonstrate that men and women differ in several behavioral preferences that critically impact team dynamics—most notably in risk aversion and competitiveness.

In a seminal, highly cited review in the *Journal of Economic Literature*, Croson and Gneezy (2009) synthesized decades of experimental evidence, concluding that women are reliably more risk-averse than men across a wide variety of contexts[4]. Furthermore, women are generally more averse to competitive environments; even when women possess the requisite skills to succeed, they are statistically less likely to self-select into tournament-style payment schemes or zero-sum competitions compared to men, often by a factor of two-to-one[6]. Croson and Gneezy also noted that women's social preferences are more "situationally specific" and malleable than men's, adapting more fluidly to the immediate social context and displaying less rigid reliance on absolute rules[6].

Distinguishing Averages from Individuals: The Overlapping Bell Curves

When translating these behavioral and physiological findings to team composition, it is vital to distinguish what is known about *statistical averages* from what can be said about *individuals*. Effect sizes such as a Cohen's *d* of 0.47 (as seen in the RMET female advantage) represent a substantial overlap between the male and female distributions. In practical terms, a *d* of 0.47 means that approximately 32% of men will score higher in social sensitivity than the average woman. Similarly, while men are more risk-tolerant on average, a significant percentage of women will exhibit higher risk tolerance than the average man.

Therefore, while a randomly selected all-female team is statistically highly likely to possess higher average social sensitivity and lower average risk tolerance than a randomly selected all-male team, predicting the dynamics of any *specific* mixed-sex team based purely on its gender ratio is inherently fallible. The overlapping distributions mean that individual variances in personality traits (like the Big Five), general cognitive ability (*g*), and domain-specific expertise frequently override statistical gender averages within small groups. Gender serves as a probabilistic proxy for cognitive diversity, but it is not a deterministic absolute.

Layer 3: Culture (Context and Variability)

The final layer is cultural, encompassing the local norms, organizational climates, and psychological environments that govern how physiological and behavioral traits manifest. A team's culture acts as an environmental moderator, either suppressing or amplifying individual biological and behavioral differences.

The most critical cultural construct in team literature is "psychological safety." Defined in a landmark 1999 study in *Administrative Science Quarterly* by Amy Edmondson, psychological safety is "a shared belief held by members of a team that the team is safe for interpersonal

risk-taking"[9]. Based on a sample of 51 work teams, Edmondson found that psychological safety mitigates the fear of embarrassment or punishment, thereby fostering learning behaviors (such as reporting errors, asking for help, or proposing unconventional ideas) which ultimately drive team performance[10].

Crucially, psychological safety is a local, cultural phenomenon, driven primarily by team leadership and immediate peer norms rather than the innate psychological traits of the members[11]. A meta-analysis by Newman et al. (2017) aggregating 136 empirical studies found psychological safety reliably predicted learning behavior, performance, and creativity, with effect sizes ranging from $r = 0.21$ to $r = 0.37$[11]. If a mixed-sex team possesses high baseline social sensitivity (Layer 1) but operates in a toxic, highly punitive organizational culture (Layer 3), the team will not actualize its collaborative potential. Conversely, a psychologically safe environment can encourage risk-averse individuals to contribute boldly and competitive individuals to collaborate harmoniously.

Layer	Domain focus	Key Construct	Primary Measurement	Meta-Analytic Support
Physiology	Neural/Cognitive	Theory of Mind (ToM)	RMET (Reading the Mind in the Eyes)	Kirkland et al., 2013 ($d = 0.47$); Greenberg et al., 2022
Behavior	Behavioral Economics	Risk & Competition	Tournament selection / Gamble tasks	Croson & Gneezy, 2009
Culture	Org. Psychology	Psychological Safety	Edmondson's 7-item survey scale	Newman et al., 2017 ($r = 0.21$ to 0.37)

The Collective Intelligence Construct: A Fierce Controversy

The intersection of individual traits, gender composition, and team performance was revolutionized by a highly influential paradigm introduced by Anita Williams Woolley, Thomas W. Malone, and colleagues. However, this paradigm has since become the center of a fierce methodological controversy, representing a critical debate in modern psychometrics.

The Proponents: Evidence for a 'c' Factor

In a 2010 paper published in *Science*, Woolley et al. proposed that groups, much like individuals, possess a stable, generalizable intelligence factor. In psychometrics, individual intelligence

across diverse cognitive tasks is largely explained by a single latent variable known as general cognitive ability, or *g*, which typically accounts for 30% to 50% of the variance in test scores and represents one of the most replicated findings in psychology[12]. Woolley et al. hypothesized the existence of a "collective intelligence" factor, termed *c*, which dictates a group's ability to perform across a wide variety of collaborative tasks, distinct from the mere sum of the members' individual IQs[12].

To test this hypothesis, they randomly assigned individuals to 40 three-person groups (Study 1) and 152 groups of varying sizes (Study 2)[12]. Groups were given a battery of tasks drawn from the McGrath Task Circumplex, encompassing visual puzzles, brainstorming, moral judgments, and complex resource negotiation[12].

The researchers reported that a single latent factor (*c*) explained 43% of the variance in group performance across these disparate tasks[14]. Remarkably, they found that a group's collective intelligence was only moderately predicted by the average or maximum individual IQ (*g*) of its members[14]. Instead, *c* was significantly correlated with three interrelated variables:

1. **Average Social Sensitivity:** Groups with higher average scores on the RMET demonstrated higher collective intelligence[14].
2. **Conversational Turn-Taking:** Groups where a few people dominated the conversation had lower collective intelligence; equal participation predicted higher *c* (r = -0.41 correlation with variance in speaking turns)[14].
3. **Proportion of Women:** Groups with more women exhibited higher collective intelligence[14].

Woolley et al. noted that the positive effect of the proportion of women was largely statistically mediated by the RMET scores. Because women generally score higher on social sensitivity, teams with more women possessed a higher aggregate capacity for reading nonverbal cues, which in turn facilitated equal turn-taking, smoother coordination, and higher *c*[14]. This finding was subsequently supported by a meta-analysis led by researchers from the same academic cluster (Riedl et al., covering 22 studies, 1,356 groups, and 5,279 individuals), which claimed to confirm the robust structure of the *c* factor (reporting an average variance extracted of 43.6%)[15]. Furthermore, Engel et al. (2014) published a replication in *PLOS One* showing that RMET scores predicted *c* equally well in online, text-based groups, suggesting that Theory of Mind abilities tap into domain-independent social reasoning beyond literal facial decoding[14].

The Critics: Methodological Flaws and Replication Failures

The concept of a universal collective intelligence factor has faced intense, high-profile criticism, most notably from psychometricians Marcus Credé and Garett Howardson. Presenting the strongest evidence on the opposing side, their 2017 paper published in the *Journal of Applied Psychology* re-examined the data from six previously published samples used to support the *c* construct, concluding that the empirical support for collective intelligence was "generally weak"[22].

Credé and Howardson raised several severe methodological critiques. First, they demonstrated that the general factor (*c*) explained a very low percentage of the variance in actual group performance. While Woolley et al. reported variance extraction of 43% using specific analytical

models, Credé and Howardson calculated the Average Variance Extracted (AVE) using the Fornell and Larcker (1981) formula across the datasets to be between 22% and 35%—well below the 50% threshold generally required in psychometrics to indicate a substantive latent factor[22]. When they applied principal axis factoring to the data, they found that a single factor had an eigenvalue of only 1.05 (accounting for 17% of variance), falling short of the standard threshold (1.38) required to declare a single dominant intelligence factor[24].

Second, Credé and Howardson argued that the modest covariation between tasks that *did* exist was likely inflated by statistical artifacts. Specifically, the nested nature of the data and the presence of low-effort responding (where a poorly motivated group performs poorly on all tasks not because they lack *c*, but because they lack effort) artificially inflated the apparent general factor[22].

Third, and most damagingly, they conducted a meta-analysis showing that when individual intelligence was rigorously controlled for via established *g*-factor tests, the incremental validity of the *c*-factor collapsed. In a meta-analysis of 857 groups across eight independent samples, they found only a weak-to-moderate correlation (r = 0.26, 95% CI [.10, .40]) between the purported *c*-factor and external group criterion tasks[20]. Furthermore, they found that the original studies often failed to properly measure or control for the *g*-factor of the individuals[25]. They cautioned the academic and business communities against prematurely abandoning the selection of high-IQ individuals in favor of optimizing for "collective intelligence" metrics[25].

The controversy remains open. While the finding that groups perform better when members possess high social sensitivity and communicate equally is generally accepted as a reliable behavioral phenomenon of group dynamics, the existence of *c* as a measurable, domain-general intelligence equivalent to *g* remains highly contested and has failed independent psychometric replication.

Gender Composition and Team Performance: The Empirical Record

Moving beyond the psychometric debates regarding latent intelligence, a vast empirical literature exists measuring the direct causal relationship between the sex composition of a team and its tangible output (e.g., sales, profits, decision quality, and innovation). The results are highly heterogeneous, heavily dependent on the communication medium, the nature of the task, and the structural dynamics of the group.

When Mixed Teams Outperform: The Case for 50/50 Diversity

Field experiments in business settings provide some of the strongest causal evidence for the efficacy of mixed-sex teams. In a highly cited 2013 study published in *Management Science*, Hoogendoorn, Oosterbeek, and van Praag conducted a randomized field experiment with approximately 550 undergraduate business students divided into 45 teams of 12 members[26]. The teams were tasked with launching and running a real business venture as part of an entrepreneurship curriculum, selling stock, electing officers, and marketing products over an extended period[26].

The researchers systematically manipulated the gender ratios of the teams. They found a

pronounced, non-linear relationship: teams with an equal gender mix (roughly 50/50) significantly outperformed male-dominated teams in terms of sales and profits[26]. Interestingly, profits increased linearly as the share of women increased up to 50%, but the relationship flattened for higher shares of women (though the data distribution at the extreme high end limited definitive conclusions about overwhelmingly female-dominated groups)[27].

Despite measuring a wide array of potential mechanisms—including conflict levels, friendship ties, decision-making processes, atmosphere, mutual monitoring, and learning—the authors found no single behavioral mediator that adequately explained the performance boost[27]. The implication is that in large groups engaged in complex, multifaceted tasks requiring prolonged real-world execution, the integration of diverse cognitive styles, life experiences, and problem-solving approaches yields a synergistic performance premium that is greater than the sum of its parts.

When Mixed Teams Underperform: Task Alignment and Communication Breakdowns

Conversely, under specific task constraints and communication mediums, gender diversity can impede team performance. The empirical record highlights two distinct scenarios where mixed-sex or all-female teams underperform relative to all-male teams: tasks rewarding extreme risk-taking, and online audio-only communication environments.

Risk, Aggressiveness, and Task Alignment: In a 2012 study in *Management Science*, Apesteguia, Azmat, and Iriberri analyzed a large-scale business strategy simulation played by undergraduate and MBA students in teams of three[28]. The simulation required teams to act as general managers making aggressive strategic market choices.

The analysis revealed that teams formed by three women were significantly outperformed by all other gender combinations[28]. Looking across the performance distribution, three-women undergraduate teams at the bottom decile were outperformed by an enormous 0.47 standard deviations from the mean, while at the top decile they trailed by 0.09 standard deviations[29]. For MBA students, the optimal team composition at the highest echelons of performance was two men and one woman[30].

The authors traced this performance gap directly to behavioral decision-making differences. Three-women teams were systematically less aggressive in their pricing strategies, invested significantly less in Research and Development (R&D), and diverted more capital toward social sustainability initiatives[28]. In an environment where the algorithmic simulation rewarded ruthless market pricing and high-risk R&D investments, the baseline behavioral tendency toward risk aversion (as identified by Croson & Gneezy) served as a direct handicap[6]. This highlights that when the specific task requires traits heavily skewed toward one end of the behavioral distribution (e.g., extreme risk tolerance), homogeneous groups possessing those traits will mathematically outperform heterogeneous groups.

Communication Mediums and Verbal Dominance: A more recent 2024 study in *Management Science* by Hardt, Mayer, and Rincke explored how communication mediums drastically alter mixed-sex dynamics[31]. Using an online experiment with 1,368 subjects in 342 four-person teams, participants worked on a gender-neutral cognitive task inside an audio-only chat room

(where visual cues were absent)[31].

The researchers found that all-male teams communicated far more extensively than mixed and all-female teams, specifically generating a higher volume of "topic words" related directly to solving the task[31]. Consequently, all-male teams outperformed both mixed and all-female teams in this specific paradigm.

Crucially, the study illuminated a striking communication breakdown within mixed-sex teams. When placed in a mixed environment, men dramatically increased their verbal output compared to when they were in all-male teams, while women adjusted their communication behavior in the opposite direction, speaking significantly less[31]. This resulted in a massive conversational gender gap, with men uttering almost 70% more words than women in mixed groups[31]. When stratifying by cognitive skill level (proxied by GPA), the researchers found that high-skilled men spoke the most, followed closely by low-skilled men. High-skilled women spoke significantly less than low-skilled men[31].

This study reveals a critical boundary condition to Woolley et al.'s findings regarding equal turn-taking[12]. In unstructured, audio-only environments lacking visual cues to trigger RMET-style social sensitivity, male verbal dominance suppresses female participation. This dynamic negates the potential benefits of cognitive diversity and renders all-male teams more efficient due to higher raw information throughput.

Summary of Experimental Configurations and Outcomes

Study	Journal & Year	Sample	Group Size	Task Environment	Optimal Performing Composition	Key Mechanism Identified
Woolley et al.	*Science* (2010)	~192 groups	3 to 5	Lab tasks (McGrath Circumplex)	High proportion of women	Social sensitivity (RMET); equal turn-taking.
Hoogendoorn et al.	*Management Science* (2013)	45 groups (N~550)	12	Real-world business venture	50/50 Equal Mix	Synergistic diversity (specific mediator elusive).
Apesteg	*Manage*	Large	3	Strategic	2 Men, 1	Higher

uia et al.	*ment Science* (2012)	cohort		business simulatio n	Woman (at top level)	R&D investme nt; aggressiv e pricing.
Hardt et al.	*Manage ment Science* (2024)	342 groups (N=1368)	4	Online audio-ch at puzzle	All-Male	High volume of topic-wo rd communi cation; male verbal dominan ce.

Corporate Boardrooms: The Quota Controversy

If experimental evidence on small teams reveals extreme context dependency, the macro-level evidence from corporate boardrooms is similarly nuanced, complicated by severe econometric identification challenges and reverse causality (do diverse boards create successful companies, or do successful companies have the luxury to prioritize diverse boards?).

In corporate governance, Adams and Ferreira (2009) demonstrated in the *Journal of Financial Economics* that female directors possess distinct behavioral patterns: they attend board meetings more regularly, are more likely to join monitoring committees, and hold CEOs more strictly accountable[32].

However, translating improved monitoring into financial performance has sparked immense academic warfare, largely centered on the Norwegian gender quota, which mandated that corporate boards comprise at least 40% women. A highly influential 2012 paper by Ahern and Dittmar in the *Quarterly Journal of Economics* reported that the forced introduction of female directors led to younger, less experienced boards, causing a significant decline in firm value (measured by Tobin's Q) and a deterioration in operating performance[33]. Similarly, Matsa and Miller (2013) in the *American Economic Journal* found that quota-affected firms saw reduced short-term profitability due to fewer workforce downsizings, leading to higher relative labor costs[33].

This narrative of quotas destroying shareholder value has been fiercely rebutted. Eckbo, Nygaard, and Thorburn (2022) completely re-analyzed Ahern and Dittmar's data in an *ECGI Working Paper*. They argued that the original authors assigned negative market reactions to the wrong event dates and failed to adjust for the contemporaneous cross-correlation of stock

returns (a fatal statistical flaw when analyzing a regulatory shock that affects all firms simultaneously)[34]. When these econometric errors were corrected, Eckbo et al. found that the valuation effect of the quota on Norwegian firms was statistically indistinguishable from zero, and the pool of qualified female directors was deep enough to avoid shareholder-borne costs[34].

Looking globally, Post and Byron (2015) conducted a massive meta-analysis of 140 studies on board gender diversity[35]. They concluded that female board representation is generally positively related to *accounting returns* (metrics like ROA and ROE, which monitor internal efficiency and operational oversight) but has a near-zero or neutral relationship with *market performance* (how external investors value the firm's future growth)[35]. The consensus view in the literature is that gender diversity improves board monitoring, ethical compliance, and ESG (Environmental, Social, and Governance) scores, but does not inherently generate windfall market profits, nor does it destroy shareholder value.

Group Size, Faultlines, and Tokenism

A critical variable explaining why mixed-sex teams succeed in some contexts (Hoogendoorn's 12-person teams) and fail in others (Apesteguia and Hardt's 3- and 4-person teams) is group size and the resulting demographic "faultlines."

Developed by Lau and Murnighan in a highly cited 1998 article in the *Academy of Management Review*, faultline theory posits that a group's demographic composition can create hypothetical dividing lines that split a team into adversarial sub-groups based on aligned individual attributes[41]. A faultline is strongest when multiple demographic attributes align perfectly. For example, in a four-person group consisting of two young female marketers and two older male engineers, the faultline is incredibly strong. Grounded in social identity theory and self-categorization theory, strong faultlines cause members to identify with their sub-group rather than the team, triggering an "us versus them" dynamic[41]. This increases relationship conflict, reduces psychological safety, and severely impairs performance[43].

Group size profoundly interacts with faultlines. In Hoogendoorn's study, the teams consisted of 12 people. A 50/50 split in a 12-person group creates a critical mass of 6 men and 6 women. This is large enough to prevent tokenism but diverse enough that gender identity is likely intersected by other differing traits (e.g., personality, nationality, skillset), which creates cross-cutting ties and *weakens* the faultline[27].

Conversely, in small micro-teams, a mixed-gender setup inherently isolates one person as a "solo" or token member. Bear and Woolley (2011) note that solo women in majority-male groups are systematically less talkative and less influential than women in balanced groups[46]. This tokenism explains why 2:1 splits in small triads often result in communication imbalances and underperformance. To maximize the collective intelligence of a mixed-sex team, organizational designers must actively manage faultlines, recognizing that diversity without critical mass often breeds isolation rather than innovation.

Summary of Robust Findings

1. Women possess a replicable, cross-cultural physiological advantage in decoding non-verbal emotional cues from the face, yielding a small-to-medium effect size (*d* =

0.47) on the RMET.
2. Men demonstrate higher baseline behavioral tolerances for financial and strategic risk, and a higher propensity to self-select into competitive, tournament-style environments.
3. These behavioral and physiological traits exist on highly overlapping normal distributions; individual variance frequently overrides statistical gender averages within specific groups.
4. The concept of a general "collective intelligence" (*c*) factor analogous to individual IQ (*g*) is highly contested, with critics demonstrating it explains far less variance (22-35%) in performance than originally claimed and lacks incremental validity over individual IQ.
5. In unstructured, audio-only online environments, mixed-sex teams experience severe conversational imbalances, with men speaking up to 70% more than women, suppressing female participation and team output.
6. In simulated business games rewarding aggressive pricing and high-risk R&D investments, all-female teams systematically underperform due to average differences in risk tolerance.
7. In large-scale, real-world operational teams (e.g., 12-person groups), an equal 50/50 gender mix optimizes sales and profits compared to male-dominated teams, suggesting synergistic diversity benefits at scale.
8. Small mixed-sex teams are highly susceptible to demographic "faultlines," where token or solo members withdraw participation, triggering communication breakdowns and in-group/out-group conflict.
9. At the macro-level, female representation on corporate boards is positively associated with internal accounting returns and monitoring oversight, but has a statistically neutral effect on market valuation (e.g., Tobin's Q).
10. Psychological safety—a cultural norm of interpersonal trust set by local leadership—acts as a master moderator, determining whether the cognitive diversity of a mixed-sex team is harnessed or suppressed.

Contested or Open Questions

- **The Validity of the *c*-Factor:** Does collective intelligence truly exist as a stable, domain-general psychometric property of groups, or is it merely a statistical artifact of nested data and low-effort responding as argued by Credé and Howardson?
- **Digital vs. Physical Collaboration:** How do the absence of visual cues in remote and digital work environments systematically alter the influence of female social sensitivity (RMET) on team coordination?
- **Critical Mass vs. Tokenism:** What is the exact mathematical threshold in varying group sizes where a minority demographic transitions from being an isolated "token" to a culturally influential block capable of shifting group norms?
- **Evolution of Risk Aversion:** To what extent are the documented gender differences in risk and competition biologically hardwired versus a product of malleable cultural socialization?
- **Non-linear Diversity Returns:** Why do the profit benefits of gender diversity in large teams appear to flatten out or yield diminishing returns once a 50/50 threshold is

crossed, rather than continuing linearly?

5 to 10 Most Important Sources to Read in Full

1. **Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010).** "Evidence for a collective intelligence factor in the performance of human groups." *Science*.
2. **Credé, M., & Howardson, G. (2017).** "The structure of group task performance: A second look at the evidence for the collective intelligence construct." *Journal of Applied Psychology*.
3. **Hoogendoorn, S., Oosterbeek, H., & van Praag, M. (2013).** "The impact of gender diversity on the performance of business teams: Evidence from a field experiment." *Management Science*.
4. **Hardt, D., Mayer, L., & Rincke, J. (2024).** "Who Does the Talking Here? The Impact of Gender Composition on Team Interactions." *Management Science*.
5. **Croson, R., & Gneezy, U. (2009).** "Gender differences in preferences." *Journal of Economic Literature*.
6. **Lau, D. C., & Murnighan, J. K. (1998).** "Demographic diversity and faultlines: The compositional dynamics of organizational groups." *Academy of Management Review*.
7. **Edmondson, A. (1999).** "Psychological safety and learning behavior in work teams." *Administrative Science Quarterly*.
8. **Greenberg, D. M., et al. (2022).** "Sex and age differences in “theory of mind” across 57 countries using the English version of the “Reading the Mind in the Eyes” Test." *PNAS*.
9. **Apesteguia, J., Azmat, G., & Iriberri, N. (2012).** "The impact of gender composition on team performance and decision making: Evidence from the field." *Management Science*.
10. **Eckbo, B. E., Nygaard, K., & Thorburn, K. S. (2022).** "Valuation effects of Norway’s board gender-quota law revisited." *ECGI Working Paper*.

Implications for Living Together

Translating psychometric data and organizational behavior into rules for daily life requires strict avoidance of the naturalistic fallacy: the philosophical error of assuming that because a behavioral difference (like male risk tolerance or female social sensitivity) is statistically "natural" or biological, it is therefore morally "good," inevitable, or dictates how any individual *should* behave. The existence of overlapping distributions ensures that averages do not define individual potential.

However, the empirical record offers profound structural insights for couples, families, and organizations. The data clearly shows that men and women bring statistically distinct cognitive and behavioral baselines to cooperative tasks. Because these distributions overlap heavily, relying on gender stereotypes to assign roles within a family or a small organizational team is both inefficient and factually inaccurate.

Instead, the true utility of understanding mixed-sex dynamics lies in mastering the concepts of *faultlines*, *critical mass*, and *communication architectures*. In households, small family businesses, or executive committees, isolated demographics (e.g., a mother surrounded by three sons, or a single female executive on a board of men) can easily be drowned out by

dominant communication styles, specifically the well-documented male propensity to occupy verbal space, particularly in unstructured environments.
If families and organizations want to harvest the superior social sensitivity, ethical monitoring, and unique risk-assessment capabilities that women, on average, possess, they cannot rely on passive inclusion. They must actively engineer *psychological safety* and enforce equal turn-taking norms to overcome natural faultlines. Ultimately, the successful coexistence of men and women relies not on ignoring our statistical differences, but on designing environments that protect minority voices, mitigate tokenism, and transform inherent cognitive friction into collective problem-solving.

Works cited

1. (PDF) Meta-analysis reveals adult female superiority in "Reading the, https://www.researchgate.net/publication/260712981_Meta-analysis_reveals_adult_female_superiority_in_Reading_the_Mind_in_the_Eyes_Test
2. Sex and age differences in 'theory of mind' across 57 countries using, https://www.repository.cam.ac.uk/bitstreams/8d5bcbc4-e83f-4a21-a83e-cde0f3970c0b/download
3. Sex and age differences in “theory of mind” across 57 countries, https://www.pnas.org/doi/abs/10.1073/pnas.2022385119?fbclid=PAAaZ1vpMMSfVkJ2nU8iNIo5wjVC61SzrNjQXY_Jb55kKgbKVtX_gYpalsE2Y
4. Behavioral Preference and Gender | PDF | Risk Aversion - Scribd, https://www.scribd.com/document/443277712/behavioral-preference-and-gender
5. Gender Differences in Preferences - ResearchGate, https://www.researchgate.net/publication/227361832_Gender_Differences_in_Preferences
6. Gender Differences in Preferences - Rady School of Management, https://rady.ucsd.edu/_files/faculty-research/uri-gneezy/gender-differences-preference.pdf
7. Gender and competition in adolescence: task matters - ZORA, https://www.zora.uzh.ch/server/api/core/bitstreams/26f292f1-66b7-4b53-a3a1-d67b46ebf726/content
8. Distributional preferences and competitive behavior - PMC - NIH, https://pmc.ncbi.nlm.nih.gov/articles/PMC3617875/
9. Why Psychological Safety Improves The Effectiveness Of Your Team, https://medium.com/the-liberators/why-psychological-safety-improves-the-effectiveness-of-your-team-7592d76f3c9b
10. Psychological Safety and Learning Behavior in Work Teams - MIT, https://web.mit.edu/curhan/www/docs/Articles/15341_Readings/Group_Performance/Edmondson%20Psychological%20safety.pdf
11. Psychological Safety: Why Teams Speak Up Without Fear, https://yukaichou.com/gamification-analysis/psychological-safety-edmondson-team-learning/
12. Evidence for a Collective Intelligence Factor in the Performance of,

https://www.researchgate.net/publication/47369848_Evidence_for_a_Collective_Intelligence_Factor_in_the_Performance_of_Human_Groups
13. Evidence for a Collective Intelligence Factor in the Performance of, https://ofew.berkeley.edu/sites/default/files/evidence_for_a_collective_intelligence_factor_in_the_performance_of_human_groups_woolley_et_al.pdf
14. Reading the Mind in the Eyes or Reading between the Lines, https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0115212
15. Quantifying collective intelligence in human groups - PNAS, https://www.pnas.org/doi/10.1073/pnas.2005737118
16. Evidence from a Collective Intelligence Factor in the Performance of, https://gap.hks.harvard.edu/evidence-collective-intelligence-factor-performance-human-groups
17. (PDF) Reading the Mind in the Eyes or Reading between the Lines, https://www.researchgate.net/publication/269773928_Reading_the_Mind_in_the_Eyes_or_Reading_between_the_Lines_Theory_of_Mind_Predicts_Collective_Intelligence_Equally_Well_Online_and_Face-To-Face
18. A recipe for dyadic collective intelligence for well-structured tasks, https://pmc.ncbi.nlm.nih.gov/articles/PMC12460217/
19. Problem Solving: When Groups Perform Better Than Teammates, https://www.jasss.org/23/3/4.html
20. High-performing teams: Is collective intelligence the answer?, https://acuresearchbank.acu.edu.au/bitstreams/a364226f-a704-451f-9f76-cfc2453c3259/download
21. Quantifying Collective Intelligence in Human Groups - INSEAD, https://sites.insead.edu/facultyresearch/research/file.cfm?fid=67158
22. (PDF) The structure of group task performance: A second look at the, https://www.researchgate.net/publication/308399872_The_structure_of_group_task_performance_A_second_look_at_the_evidence_for_the_collective_intelligence_construct
23. The structure of group task performance-A second look at "collective, https://pubmed.ncbi.nlm.nih.gov/29016184/
24. Clarifying the Structure of Collective Intelligence in Teams: A Meta, https://www.researchgate.net/publication/336103287_Clarifying_the_Structure_of_Collective_Intelligence_in_Teams_A_Meta-Analysis
25. g versus c: comparing individual and collective intelligence across, https://pmc.ncbi.nlm.nih.gov/articles/PMC8019454/
26. The Impact of Gender Diversity on the Performance of Business Teams, https://www.researchgate.net/publication/228271389_The_Impact_of_Gender_Diversity_on_the_Performance_of_Business_Teams_Evidence_from_a_Field_Experiment
27. The Impact of Gender Diversity on the Performance of Business Teams, https://gap.hks.harvard.edu/impact-gender-diversity-performance-business-teams-evidence-field-experiment
28. (PDF) The Impact of Gender Composition on Team Performance, https://www.researchgate.net/publication/228516728_The_Impact_of_Gender_Co

mposition_on_Team_Performance_and_Decision_Making_Evidence_from_the_Field
29. The Impact of Gender Composition on Team Performance and, https://www.researchgate.net/publication/261970788_The_Impact_of_Gender_Composition_on_Team_Performance_and_Decision_Making_Evidence_from_the_Field
30. MANAGEMENT SCIENCE - Jose Apesteguia, https://jose-apesteguia.github.io/Gender_Teams.pdf
31. Who Does the Talking Here? The Impact of Gender Composition on, https://pubsonline.informs.org/doi/10.1287/mnsc.2023.03411
32. Do Female Directors Boost Firm Performance?, https://www.oenb.at/dam/jcr:632ea176-2b03-47af-a705-bdad1e11a1a6/Firm_Performance_Gender_2025.pdf
33. Gender diversity in corporate boards: Evidence from quota-implied, https://www.aeaweb.org/conference/2020/preliminary/paper/Q8iNE4eK
34. Valuation effects of Norway's board gender-quota law revisited, https://www.ecgi.global/sites/default/files/working_papers/documents/eckbonygaardthorburnfinal.pdf
35. Gender Diversity on Corporate Boards and Firm Performance, https://www.preprints.org/manuscript/202503.2128
36. How costly is forced gender-balancing of corporate boards?, https://www.ecgi.global/sites/default/files/working_papers/documents/SSRN-id2746786.pdf
37. Can Female Directors Shrink the Gender Gap? Evidence from France, https://thedocs.worldbank.org/en/doc/9fe224595c1ecfe9e508d7f3be205a8d-0080012025/related/S2-3-P-women-on-board.pdf
38. Does Gender Diversity on Boards Really Boost Company, https://knowledge.wharton.upenn.edu/article/will-gender-diversity-boards-really-boost-company-performance/
39. Gender diversity in the board, women's leadership and business, https://upcommons.upc.edu/bitstreams/47cf4d61-8937-4dd7-ac07-41a9725876bc/download
40. Does Gender Matter? Female Representation on Corporate Boards, https://pmc.ncbi.nlm.nih.gov/articles/PMC4473005/
41. Subgroup Splits in Diverse Work Teams - Frontiers, https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2021.595720/full
42. How Surface-Level and Deep-Level Faultlines Influence Team, https://www.cambridge.org/core/journals/management-and-organization-review/article/how-surfacelevel-and-deeplevel-faultlines-influence-team-performance-through-subgroup-formation-and-team-interaction-quality-a-metaanalytic-review/D3EA57D399DA0CF58D16B0A58BB22745
43. Research on Team Faultlines: A Literature Review - SciRP.org, https://www.scirp.org/journal/paperinformation?paperid=91241
44. (PDF) Interactions Within Groups and Subgroups: The Effects of,

https://www.researchgate.net/publication/259461843_Interactions_Within_Groups_and_Subgroups_The_Effects_of_Demographic_Faultlines
45. Demographic faultlines and creativity in diverse groups, https://www.emerald.com/books/edited-volume/13045/chapter/83466992/Demographic-faultlines-and-creativity-in-diverse
46. (PDF) The role of gender in team collaboration and performance, https://www.researchgate.net/publication/228196582_The_role_of_gender_in_team_collaboration_and_performance
47. Revisiting the Effects of Gender Diversity in Small Groups on ... - PMC, https://pmc.ncbi.nlm.nih.gov/articles/PMC8543034/