Hormone-Diversity Fit: Collective Testosterone Moderates the Effect of Diversity on Group Performance

Journal:	*Psychological Science*
Manuscript ID	PSCI-17-0205.R3
Manuscript Type:	Research Article
Date Submitted by the Author:	03-Oct-2017
Complete List of Authors:	Akinola, Modupe; Columbia University, Business School Page-Gould, Elizabeth; University of Toronto, Psychology Mehta, Pranjal; University of Oregon, Psychology Liu, Zaijia; Columbia University, Business School
Keywords:	Testosterone, Status, Groups, Diversity, Performance

RUNNING HEAD: Hormone-Diversity Fit

Hormone-Diversity Fit:

Collective Testosterone Moderates the Effect of Diversity on Group Performance

Modupe Akinola[a,*]

Elizabeth Page-Gould[b]

Pranjal H. Mehta[c]

Zaijia Liu[a]

[a] Columbia Business School, Columbia University, New York, NY 10027; [b] Department of Psychology, University of Toronto, 100 St. George St 4th Floor, Toronto, ON, M5S 3G3, Canada, [c] Department of Experimental Psychology, University College London, 26 Bedford Way, London WC1H 0AP, United Kingdom

[*] To whom correspondence should be addressed: Modupe Akinola, Columbia University, 3022 Broadway, Uris Hall 708, New York, NY 10027. Phone: (212) 854-0756. Fax: (212) 854-3778. E-mail: makinola@columbia.edu

Abstract

Prior research has found inconsistent effects of diversity on group performance. The present research identifies hormonal factors as a critical moderator of the diversity-performance connection. Integrating the diversity, status, and hormone literatures, we predicted that groups collectively low in testosterone, which orients individuals less toward status competitions and more toward cooperation, would excel with greater group diversity. In contrast, groups collectively high in testosterone, which is associated with a heightened status drive, would be derailed by diversity. Analysis of 74 randomly assigned groups engaged in a group decision-making exercise provided support for these hypotheses. The findings suggest that diversity is beneficial for performance, but only if group-level testosterone is low; diversity has a negative effect on performance if group-level testosterone is high. Too much collective testosterone maximizes the pains and minimizes the gains from diversity.

Keywords: Diversity, testosterone, status, groups, performance

For decades, researchers have investigated the effects of diversity on group dynamics, but the nature of diversity's influence on group performance remains unclear (Jehn, Northcraft, & Neale, 1999; Mannix & Neale, 2005). On the one hand, diversity often enhances group performance as the diverging perspectives of group members can lead to better decisions and more creative ideas and solutions. On the other hand, it can also hinder performance by increasing conflict between group members (see Galinsky et al., 2015 for a review).

Diversity is particularly relevant in the context of group competition. Groups can win competitions through two routes: a) by perfecting *intragroup* processes, such as coordination and integration or b) by maximizing *intergroup* competitive motivation (Galinsky & Schweitzer, 2015). The present research examines the interplay between diversity and hormonal factors in determining group performance.

There is evidence that diverse groups tend to focus their attention on intragroup dynamics relative to homogeneous groups, often leading to greater conflict, less cohesion, and less trust across group members, all of which can undermine group performance (Kirkman, Tesluk, & Rosen, 2004; Mannix & Neale, 2005; van Knippenberg & Schippers, 2007). These findings are consistent with social identification and self-categorization theories, which suggest that diversity within a group leads group members to categorize themselves along prominent social dimensions, such as race and gender, and exaggerates the differences between group members (Tajfel & Turner, 1986). These processes can increase stereotyping (Chatman, Polzer, Barsade, & Neale, 1998), heightening group members' sensitivity to how their behavior is perceived by other group members who differ demographically (Blascovich, Mendes, & Seery, 2002).

However, this focus on intragroup differences can also be beneficial for diverse groups, serving as a catalyst for group members to consider and incorporate the potentially diverging perspectives of demographically different group members into the group process (Galinsky et al., 2015; Phillips, Mannix, Neale, & Gruenfeld, 2004). Thus, among diverse groups, a focus on intragroup dynamics can have both positive and negative effects on group performance.

In contrast to diverse groups, homogeneous groups tend to focus their attention away from intragroup dynamics and toward intergroup goals. Consistent with social identity theory, during intergroup competition, groups are generally motivated to achieve higher social standing relative to other groups, which drives group members to sacrifice individual gains in an effort to accomplish the group goal of outcompeting other groups (Hogg & Terry, 2000; Tajfel & Turner, 1986). This focus on outcompeting other groups can enhance group performance, especially when the competition is intense (Cox, Lobel, & McLeod, 1991; Murray, 1989). However, this intergroup focus could impair performance by increasing conformity pressures and stifling different perspectives from emerging within the group. Homogeneous group are particularly susceptible to conformity pressures as homogeneity can motivate a need for cohesion. For example, homogeneity can increase group members' propensity to conform to clearly inferior decisions (Gaither, Apfelbaum, Birnbaum, Babbitt, & Sommers, in press). Further, homogeneous groups can be less accurate in information processing and can lack objectivity in decision making due in part to an avoidance of disagreement, relative to diverse groups (Phillips & Apfelbaum, 2012; Sommers, 2006).

Taken together, diversity and homogeneity can each be helpful and harmful to group performance. Diverse groups have the potential to capitalize on novel perspectives but are

prone to conflict; thus, they may lack the intragroup cohesion necessary to take advantage of the diverse perspectives offered. Homogeneity solves the conflict problem but makes groups susceptible to conformity pressures that can negatively influence group performance. We help reconcile these contradictory findings by examining a critical and overlooked factor in determining whether diversity and homogeneity hurt or help group performance: The hormonal make-up of group members.

Testosterone, a steroid hormone released as the end product of the hypothalamic-pituitary-gonadal axis, is associated with greater motivation to attain status and thus is particularly relevant in competitive contexts (Mazur & Booth, 1998). High-testosterone individuals tend to outperform others in competition, exhibiting dominance-related behaviors (Coates & Herbert, 2008; Mazur & Booth, 1998). Yet in the context of groups, too much testosterone can hinder performance by creating intragroup status conflict (Mehta, Lawless, van Vugt, & Josephs, 2017; Ronay, Greenaway, Anicich, & Galinsky, 2012). In contrast, low testosterone increases the motivation to cooperate and decreases status striving (Josephs, Sellers, Newman, & Mehta, 2006; Mehta, Wuehrmann, & Josephs, 2009; Wright et al., 2012). As a result, people with low testosterone perform especially well in settings that incentivize cooperation, but they perform poorly in settings in which the focus is on outcompeting others.

Building on these separate lines of research on diversity, status, and hormones, we propose that the effect of diversity on performance will depend on a group's collective testosterone levels. According to our theoretical model of hormone-diversity fit (Fig. 1), groups collectively high in testosterone will perform optimally when group diversity is low because the lack of diversity will allow these groups to focus their competitive attention on *intergroup*

status dynamics (i.e. the motivation to outcompete other groups) but their status drive will also prevent conformity pressures. In contrast, we propose that groups collectively high in testosterone would perform poorly when group diversity is high because diversity will lead these groups to focus their attention on *intragroup* status dynamics (i.e., the motivation to outcompete other individuals within the group), leading to heightened conflict among group members. For groups collectively low in testosterone (see top row of Fig. 1), we propose that they will perform optimally when diversity is high because their cooperative focus will create the cohesion often missing from diverse groups. To summarize, our theory of hormone-diversity fit proposes that diversity will boost performance among groups collectively low in testosterone, but harm performance among groups collectively high in testosterone.

Fig. 1.
Theoretical Model of Hormone-Diversity Fit

		Diversity	
		Low Diversity Focus on *intergroup* competition	**High Diversity** Focus on *intragroup* processes
Hormone	**Low Group-Level Testosterone** Low status attainment motivation; High cooperation motivation	**Poor group performance** Sufficient intragroup cohesion but too little intergroup competitive drive to attain higher status over other groups	**Optimal group performance** Sufficient intragroup cohesion necessary to take advantage of diversity
	High Group-Level Testosterone High status attainment motivation	**Optimal group performance** Sufficient intragroup cohesion and sufficient intergroup competitive drive to attain higher status over other groups	**Poor group performance** Too much intragroup conflict due to the combination of diversity and the motive to attain higher status over other group members

The present research provides an initial test of our theory of hormone-diversity fit. Our study was designed to test the primary phenomenon that the model proposes, which is an interaction between collective hormone levels and diversity in determining group performance. However, we leave an investigation of the processes outlined in our model for follow-up research. We examine our hypothesis that group-level testosterone moderates the effect of diversity on group performance by randomly assigning individuals to groups and using a statistical methodology that takes into consideration diversity on multiple categories of difference across group members. Specifically, rather than purely measuring one dimension of group member diversity (e.g., ethnicity), we employ a faultline framework (Lau & Murnighan, 1998; Zanutto, Bezrukova, & Jehn, 2011) that examines the interaction of multiple attributes of group members and its effect on group performance while taking into consideration the collective hormonal profile of group members.

Method[1]

Participants were 370 Master in Business Administration students (mean age=27.5 years, SD=1.93; 64.1% male, 35.9% female) enrolled in both a leadership and an operations management course at Columbia Business School. The sample size was determined by the overall size of the class and the willingness of students to participate. The ethnic composition of our sample was diverse: 54.9% White, 16.5% Asian, 10.8% Hispanic, 9.5% South Asian, 4.6% Black, 1.4% South East Asian, and 2.4% other. Participants were randomly assigned to 74 groups that ranged in size from three to six people. All procedures were approved by the Columbia

[1] We tested our predictions by presenting new analyses of data from an experiment previously described by Akinola and colleagues (Akinola, Page-Gould, Mehta, & Lu, 2016).

University Institutional Review Board. The data and analysis syntax for R 3.3.0 (R Core Team, 2016) are provided on the Open Science Framework: http://osf.io/8eqtc.

One week prior to engaging in the group decision-making exercise, participants provided a saliva sample, later assayed for testosterone[2] (Salimetrics, CA). Average intra- and inter-assay coefficients of variation were 2.5% and 5.6%, respectively. Testosterone values were log-transformed prior to analysis and centered around the grand mean. Unbiased mean levels of testosterone were calculated for each group (Croon & van Veldhoven, 2007). We chose unbiased mean levels of testosterone to capture collective hormonal profiles as the average can be considered the central tendency of normally distributed variables. We also wanted to capture the testosterone levels of *all* group members, which we were best able to do by examining the group mean. However, we also conducted exploratory analyses using testosterone standard deviation, minimum, and maximum.

Diversity was computed by using group faultline analysis (Lau & Murnighan, 1998), which examines how group members differ across multiple attributes (Lau & Murnighan, 1998; Zanutto et al., 2011). Faultline analysis often offers more explanatory power than examining single-issue demographic characteristics (Lau & Murnighan, 2005). To illustrate our faultline approach to computing diversity, Table 1 highlights the degree of diversity of five groups and categorizes these groups by high and low diversity. For instance, the group in our sample with the lowest diversity was a five-person group consisting of three White males from the US and two White females, one of whom was from the US and the other from Eastern Europe (see

[2] We also measured cortisol and dehydroepiandrosterone. While neither of these hormones is the theoretical focus of the current research, we report on cortisol given extensive work on the dual hormone hypothesis (see Supporting Information).

Table 1, Group 1). This group is the least diverse with regard to ethnicity, gender, and country of origin relative to other groups. The group with the greatest diversity was a six-person group consisting of four White males, each from different countries, one Hispanic male from yet another country, and one White female whose country also differed from the five males (see Table 1, Group 4). This group can therefore be considered very diverse.

Table 1.

Examples of groups with low diversity and high diversity

Group Number	Member A	Member B	Member C	Member D	Member E	Member F	Diversity[a,b]	Fau
1	Male	Male	Male	Male	Female	Female	Low (2 align, 2 ways)	.75
	White	White	White	White	White	White		
	USA	USA	USA	USA	Bulgaria	USA		
2	Male	Male	Male	Female	Female	.	High (1 align, 1 way)	.47
	White	Asian	Indian	White	Hispanic	.		
	USA	Korea	USA	USA	USA	.		
3	Male	Male	Male	Female	Female	.	High (1 align, 2 ways)	.37
	White	Asian	White	White	Asian	.		
	USA	China	USA	USA	Japan	.		
4	Male	Male	Male	Male	Male	Female	High (0 align, 0 ways)	.35
	White	White	White	White	Hispanic	White		
	USA	Germany	Russia	USA	Brazil	Italy		

[a] Diversity is calculated using the faultline approach which focuses on the number of demographic characteristics that are aligned in the group (denoted as "align") and the possible ways in which the group can be divided based on these demographic characteristics (denoted as "ways") with the number of characteristics per group fixed at three (ethnicity, gender, and country of origin)

[b] We classified diversity based on the maximum number of aligned characteristics: high=0 or 1 aligned characteristics, low=2 or more aligned characteristics

The three demographic characteristics used in this study to calculate diversity using the faultline approach were ethnicity, gender, and country of origin. For ethnicity, 2.7% of groups were mono-ethnic, 23.0% had two ethnicities, 45.9% had three ethnicities, 27.0% had four ethnicities, and 1.4% had five different ethnicities. With regard to gender, 1.4% of the groups

had no women, 18.9% of the groups had one woman, and the remaining 79.7% had two women. Finally, for country of origin, 6.8% of groups represented five countries, 18.9% of groups represented four countries, 54.1% of groups represented three countries, 18.9% of groups represented two countries, and 1.4% of groups were all from the same country. Diversity was calculated using the equation below (Zanutto et al., 2011) using the asw.cluster package for R (Meyer & Glenz, 2013). According to Zanutto and colleagues (2011), the first step is to calculate:

$$Fau_g = \left[\frac{\sum_{j=1}^{p} \sum_{k=1}^{2} n_k^g \ (\bar{x}_{.jk} - \bar{x}_{.j\ .})^2}{\sum_{j=1}^{p} \sum_{k=1}^{2} \sum_{i=1}^{n_k^g} (\bar{x}_{ijk} - \bar{x}_{.j\ .})^2} \right] \quad g = 1,2, \ldots S,$$

where x_{ijk} is the value of the j^{th} characteristic of the i^{th} member of subgroup k, $\bar{x}_{.\ j\ .}$ is the overall group mean of characteristic j, $\bar{x}_{\ .jk}$ is the mean of characteristic j in subgroup k, and n_k^g is the number of members of the k^{th} subgroup (k = 1, 2) under split g. The second step is to calculate the maximum value of Fau_g over all possible splits g= 1, 2, . . .S (or, to avoid splits involving a subgroup consisting of a single member, we can maximize over all splits where each subgroup contains at least two members). (Zanutto et al., 2011, p. 706)

Fau is always less than or equal to one but larger than zero. The higher the value of *Fau,* the less diverse the group is as the group has many characteristics that are aligned. In our sample, the mean diversity (i.e., Fau) score across groups was .48 (*SD*=.09; range= .35-.75).

Groups engaged in an interdependent week-long computerized decision-making exercise (Littlefield Labs, Responsive Learning Technologies) simulating the supply chain process

of blood testing laboratories. Groups were employees at the blood testing laboratory responsible for managing several aspects of the lab with the goal of maximizing performance relative to other groups in the class. Each group had the responsibility of managing one laboratory outside of class time over seven days. On average, groups spent 20 to 30 hours on the group decision-making task over the course of the seven days. The task was interdependent as groups were encouraged to involve all group members in both developing and executing a strategy that would maximize the performance of the laboratory. To this end, groups made decisions together, either in person or via email, and would decide which group member would physically execute the strategy (i.e., by logging into the simulation platform and implementing the decided upon strategy) on a given day. In most cases, the responsibility for physically executing the strategy rotated across group members. Importantly, no unilateral strategic decisions were made without there being collective agreement across group members.

Group performance on day seven of the simulation (simulating 315 days of laboratory operations) was our key dependent variable. We selected performance on day seven as the key dependent variable because we wanted to understand the interplay of diversity and testosterone on the outcome that ultimately determined group status; groups were competing to win the exercise as determined by their day seven performance, which had implications for their grades and status in the class. However, for 52 of the 74 groups, we also captured performance on day five of the exercise (simulating 170 days of laboratory operations), which allowed us to conduct exploratory analyses to examine the stability of our predicted effect (see Supporting Information and Table S4). Group performance was a composite of the following measures: profitability, number of contracts, number of reorders on existing contracts, and

group rank relative to other groups. These measures were standardized and then averaged to create the aggregated group performance metric (α = 0.86).

Results

We conducted a micro-macro multilevel analysis (Croon & van Veldhoven, 2007) that modeled group performance as a function of an unbiased group mean for testosterone, group diversity, and the interaction between group diversity and group testosterone. Groups differed in the time of day of saliva collection and in size, however neither of these variables moderate our effects so we included them as covariates. We also controlled for the percentage of females in each group given that testosterone levels differ reliably between men and women. All predictors were mean centered prior to analysis.

We had nested data (i.e., individuals nested within groups), for which multilevel modeling (MLM) is a proper analysis as it accounts for the dependence of individuals within the same group. However, MLM is traditionally used to model dependent variables at the individual level, whereas our dependent variable, group performance, was measured at the group level. We therefore employed the micro-macro MLM method (Croon and van Veldhoven, 2007), which we implemented using the MicroMacroMultilevel package (Lu, Page-Gould, & Xu, 2017) in R version 3.3.3 (R Core Team, 2016). The micro-macro method views group-level testosterone as a latent variable of which the individual testosterone values are assumed to be manifestations[3]. Once the unbiased means are estimated, then they can be used in a linear regression with other group-level variables. If groups are different sizes, as our groups were,

[3] A similar approach to estimating unbiased means as the Croon and van Veldhoven (2007) method is the estimation of *Empirical Bayes Estimates* (Efron, 1975; Greenland, 2000), which yielded almost identical results as using the Croon and van Veldhoven (2007) approach. Additionally, our results remain significant when we use observed means.

the micro-macro method additionally requires that the standard errors of the slopes are corrected in the final linear regression. Additionally, we estimated effect size by converting the slope statistics into partial R^2 (Edwards, Muller, Wolfinger, Qaqish, & Schabenberger, 2008).

As predicted, the interaction between group testosterone and group diversity was significant, $b = 19.75$, $SE = 3.22$, $t(67) = 6.14$, $p < .01$, $R^2 = .36$ (Table 2).

Table 2.

Multilevel model predicting group performance

	Slope	*SE*	*Df*	*t*	*p*	R^2
Intercept	0.06	0.08	67	0.81	0.42	0.01
Time of Day	0.06	0.04	67	1.41	0.16	0.03
Group Size	0.24	0.09	67	2.75	0.01	0.10
Percent Female	-0.86	0.92	67	-0.94	0.35	0.01
Testosterone	0.01	0.41	67	0.02	0.98	0.00
Diversity	0.89	0.93	67	0.95	0.35	0.01
Diversity X Testosterone	19.75	3.22	67	6.14	<0.01	0.36

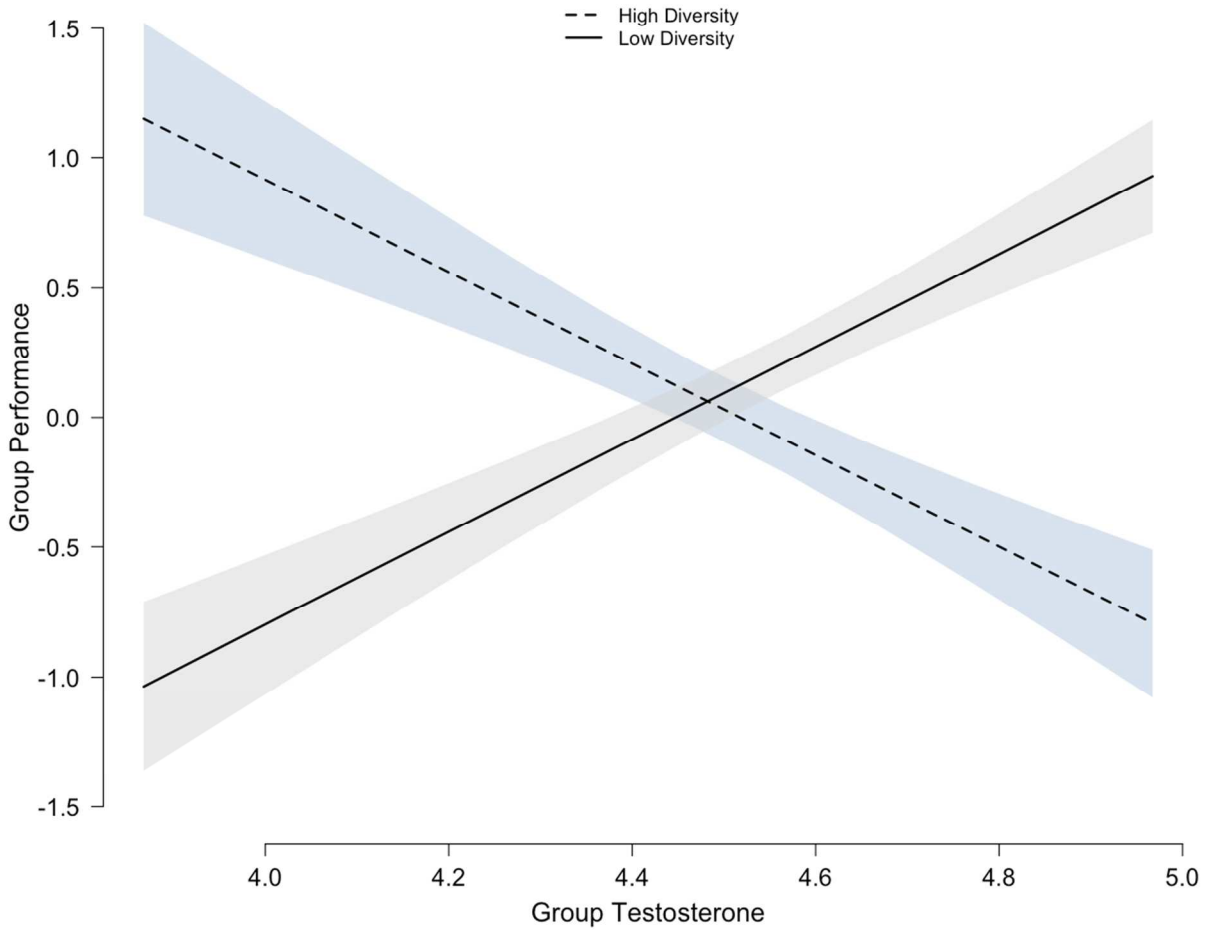
NOTE: N = 74 groups for final performance measured on day seven. Diversity was calculated using faultline analysis (Zanutto et al., 2011). Higher numbers denote lower diversity in the group as the group has many characteristics that are aligned.

Consistent with our hypothesis, when group diversity was low (Fau score was 1 SD above the mean), group testosterone significantly positively predicted performance ($b = 1.79$, $SE = 0.45$, $t(67) = 3.95$, $p < .01$, $R^2 = .19$; solid line in Fig. 2). That is, groups that were collectively high in testosterone outperformed groups collectively low in testosterone when group members had greater alignment in ethnicity, gender, and country of origin. However, when group diversity

was relatively high (Fau score was 1 SD below the mean), group testosterone significantly negatively predicted performance (b = -1.77, SE = 0.55, $t(67)$ = -3.21, $p < .01$, R^2 = .13; dashed line in Fig. 2).

Fig. 2.

Group performance as a function of group testosterone and group diversity



In other words, groups that were collectively low in testosterone outperformed groups collectively high in testosterone when group members were less aligned with regard to ethnicity, gender, and country of origin. Importantly, we observed no significant effects when examining the interaction between testosterone and ethnicity alone (b = -1.34, SE = 0.69, $t(67)$

= -1.94, p = .06, R^2 = .05), gender alone (b = 9.53, SE = 6.19, t(68) = 1.54, p = .13, R^2 = .03)[4], or country of origin alone (b = -0.91, SE = 0.86, t(67) = -1.06, p = .29, R^2 = .02).[5] These findings are consistent with research demonstrating that faultline analysis can have more explanatory power than single-issue demographic characteristics (Lau & Murnighan, 2005).

Further, to ensure that we properly controlled for gender, we also ran our analyses with log testosterone values standardized within gender as our testosterone measure. We observe the same pattern of results: the interaction between group testosterone and group diversity was significant, b = 8.82, SE = 2.18, t(67) = 4.05, p < .01, R^2 = .20 (see Supporting Information Table S3). The same analysis without controlling for the percentage of females in each group yielded a similarly significant interaction between group testosterone and group diversity, b = 9.27, SE = 2.21, t(68) = 4.19, p < .01, R^2 = .20. Additionally, we calculated a diversity score removing gender and solely including ethnicity and nationality. Again, we observed a significant interaction between group testosterone and group diversity (excluding gender), b = 8.77, SE = 2.19, t(67) = 4.01, p < .01, R^2 = .19. We re-ran this same analysis controlling for the diversity (excluding gender) x percentage of females in each group interaction and the interaction between group testosterone and group diversity (excluding gender) remained significant, b = 7.96, SE = 2.18, t(66) = 3.64, p < .01, R^2 = .17 (see Supporting Information Table S4). Taken together, these results demonstrate the robustness of our effect when taking gender into account in multiple ways.

[4] We also found no significant interactions between gender and testosterone SD, minimum, or maximum (all ps > .23).

[5] While we did observe a significant main effect of group size on performance, this effect was not consistent across all of our analyses (see Supporting Information) and therefore difficult to interpret, aligned with prior research and studies showing inconsistent effects of group size on performance (Akinola et al., 2016; Gooding & Wagner, 1985; Mao, Mason, Suri, & Watts, 2016; Wheelan, 2009).

We also repeated our primary analysis using testosterone SD, minimum, and maximum in our model. The interaction between group testosterone SD and group diversity was not significant, b = -2.06, SE = 6.20, $t(67)$ = -.33 , p = .74, R^2 < .01. However, we did observe a significant interaction using group minimum testosterone, b = 6.99, SE = 3.16, $t(67)$ = 2.21 , p = .03, R^2 = .07 and group maximum testosterone, b = 10.79, SE = 3.44, $t(67)$ = 3.13, p < .01, R^2 = .13. Importantly, when we included unbiased average group levels of testosterone, as well as minimum, and maximum testosterone and their interactions with diversity into our model, only the interaction between mean group levels of testosterone and diversity remained a reliable predictor of group performance (Table S2). Furthermore, a Bayesian model comparison (Raftery, 1995; see Supporting Information for details) suggested there was strong evidence for using the unbiased mean of testosterone over the alternative quantifications tested.

Discussion

Our findings provide preliminary support for our theoretical model of hormone-diversity fit presented in Fig. 1. We demonstrate that groups collectively high in testosterone perform optimally when group diversity is relatively low. Low diversity may allow high-testosterone groups to focus their status attainment motivations toward *outcompeting other groups*, facilitating overall group performance. In contrast, high diversity may lead groups collectively high in testosterone to focus their status attainment motives toward *outcompeting other individuals* within the group, creating intragroup conflict that undermines group performance.

Conversely, we also found that groups collectively low in testosterone performed *better* when diversity was high. Groups low in collective testosterone may experience greater intragroup cohesion as a result of the motive to cooperate (Josephs et al., 2006; Mehta et al.,

2009; Wright et al., 2012). Thus, when diversity is high, the dissimilar identities among group members may allow the group to focus attention on cooperative intragroup processes, leading to greater intragroup cohesion and better group performance. This finding is aligned with studies demonstrating that the disruptive effects of diversity can be eliminated when members of diverse groups focus on collective goals, for instance by having a culture that emphasizes collectivism, or when the task requires interdependence (Chatman, Sherman, & Doerr, 2015; Jehn et al., 1999). Importantly, our study design included random assignment of individuals to groups making it clear that our results are not due to self-sorting into groups (e.g., based on diversity dimensions). Further, the moderating effect of collective testosterone on the diversity-performance relationship could not be explained by gender differences in testosterone levels; our results remained robust using multiple ways to account for gender.

Interestingly, we found similar effects using testosterone minimum and maximum, but these effects were no longer significant when including mean testosterone levels in the model. However, since mean testosterone was significantly correlated with minimum and maximum testosterone (see Supporting Information) these findings suggest that these three different quantifications of collective hormonal profiles likely reflect similar psychological processes at play in groups. Since we did not include any intra- or inter-group process variables in this study, future research can build upon these findings and our theorizing by incorporating process measures to more directly test the predictions highlighted in our hormone-diversity fit model. Specifically, process measures that capture group cohesion and cooperation would seem especially relevant as cohesion and cooperation can mitigate the negative effects of diversity on

group performance and can enhance performance in homogeneous groups (Chatman et al., 2015; Jehn et al., 1999).

Additionally, future research is needed that examines the emergent process through which group-level testosterone and diversity affect performance by examining multiple days of performance on group decision-making tasks. While our finding that time of performance (examining both days five and seven) did not moderate our effects suggests that performance may have been stable towards the end of the task (see Supporting Information), it is possible that group performance may have shifted over the course of the week. Our theoretical model predicts that groups collectively low in testosterone but high in diversity perform well as their cooperative focus creates the cohesion. Since it can take time for groups to become cohesive (Jehn et al., 1999; Watson, Kumar, & Michaelsen, 1993), it is possible that these groups may have performed poorly at the beginning of the week but gained momentum, outperforming other groups as the week progressed. Conversely, our theory would predict that high testosterone, high diversity groups may have performed well at the beginning of the week due to status attainment motivations, but may have experienced decrements in performance over the course of the week due to intragroup competition stemming from diversity. Further exploration of these potential time of performance effects is an important avenue for future research.

Our research also demonstrates that the configuration of group members' characteristics along multiple attributes can be an even stronger determinant of group performance than individual characteristics alone. Diversity is not a unitary construct, but rather an intersection of identities (Gopaldas, 2013). By incorporating this intersectionality

perspective into research on diversity, we contribute to theory by considering the impact of different social category configurations on group performance.

In sum, by demonstrating that collective hormonal profiles implicated in status attainment and cooperation motivations moderate the effect of diversity on group performance, we open up new avenues for research on biological factors that help explain how configurations of diversity can differentially impact group performance. At the same time, we acknowledge that the current research provides only initial support for the proposed model of hormone-diversity fit. We encourage replications and new studies that explore group process-related mechanisms.

Author Contributions

M. Akinola designed the study and collected the data. Analyses were performed by E. Page-Gould. M. Akinola, Z. Liu, P.H. Mehta, and E. Page-Gould wrote the manuscript.

Acknowledgements

We thank A.D. Galinsky and W.B. Mendes for their valuable feedback.

Declaration of Conflicting Interests

The authors declare that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

This research was funded in part by the Eugene M. Lang Fund to M. Akinola and the Canada Research Chairs program, the Social Sciences and Humanities Research Council of Canada, and an Early Researcher Award from the Ontario Ministry of Research and Innovation to E. Page-Gould. P.H. Mehta was supported by National Science Foundation Grant #1451848.

References

Akinola, M., Page-Gould, E., Mehta, P. H., & Lu, J. G. (2016). Collective hormonal profiles predict group performance. *Proceedings of the National Academy of Sciences, 113* (35), 9774-9779.

Blascovich, J., Mendes, W. B., & Seery, M. D. (2002). Intergroup encounters and threat: A multi-method approach. In D. Mackie & E. Smith (Eds.), *From prejudice to intergroup emotions: Differentiated reactions to social groups* (pp. 89–110). New York: Psychology Press.

Chatman, J. A., Polzer, J. T., Barsade, S. G., & Neale, M. A. (1998). Being different yet feeling similar: The influence of demographic composition and organizational culture on work processes and outcomes. *Administrative Science Quarterly*, *41*, 749-780.

Chatman, J. A., Sherman, E. L., & Doerr, B. M. (2015). Making the Most of Diversity: How Collectivism Mutes the Disruptive Effects of Demographic Heterogeneity on Group Performance

Coates, J. M., & Herbert, J. (2008). Endogenous steroids and financial risk taking on a London trading floor. *Proceedings of the National Academy of Sciences, 105*(16), 6167-6172.

Cox, T. H., Lobel, S. A., & McLeod, P. L. (1991). Effects of ethnic group cultural differences on cooperative and competitive behavior on a group task. *Academy of Management Journal,* 34(4), 827-847.

Croon, M. A., & van Veldhoven, M. J. (2007). Predicting group-level outcome variables from variables measured at the individual level: A latent variable multilevel model. *Psychological Methods, 12*(1), 45-57.

Edwards, L. J., Muller, K. E., Wolfinger, R. D., Qaqish, B. F., & Schabenberger, O. (2008). An R2 statistic for fixed effects in the linear mixed model. *Statistics in Medicine, 27*, 6137-6157.

Efron, B. (1975). Biased versus unbiased estimation. *Advances in Mathematics*, 16, 259-277.

Gaither, S. E., Apfelbaum, E. P., Birnbaum, H. J., Babbitt, L. G., & Sommers, S. R. (In Press). Mere membership in racially diverse groups reduces conformity. *Social Psychological and Personality Science.*

Galinksky, A. D., Schweitzer, M. (2015). Friend and foe: When to cooperate, when to compete, and how to succeed at both. New York: Crown Business.

Galinsky, A. D., Todd, A. R., Homan, A. C., Phillips, K. W., Apfelbaum, E. P., Sasaki, S. J., ... & Maddux, W. W. (2015). Maximizing the gains and minimizing the pains of diversity: A policy perspective. *Perspectives on Psychological Science,* 10(6), 742-748.

Gooding, R.Z., Wagner, J.A. III. (1985). A meta-analytic review of the relationship between size and performance: The productivity and efficiency of organizations and their subunits. *Administrative Science Quarterly, 30* (4), 462–481

Gopaldas, A. (2013). Intersectionality 101. *Journal of Public Policy & Marketing*, *32*, 90-94.

Greenland, S. (2000). Principles of multilevel modelling. *International Journal of Epidemiology, 29*, 158-167.

Hogg, M. A., & Terry, D. I. (2000). Social identity and self-categorization processes in organizational contexts. *Academy of Management Review, 25*(1), 121-140

Jehn, K. A., Northcraft, G. B., & Neale, M. A. (1999). Why differences make a difference: A field study of diversity, conflict and performance in workgroups. *Administrative Science Quarterly, 44*(4), 741-763.

Josephs, R. A., Sellers, J. G., Newman, M. L., & Mehta, P. H. (2006). The mismatch effect: when testosterone and status are at odds. *Journal of Personality and Social Psychology, 90*(6), 999-1013.

Kirkman, B. L., Tesluk, P. E., & Rosen, B. (2004). The impact of demographic heterogeneity and team leader-team member demographic fit on team empowerment and effectiveness. *Group and Organization Management, 29*(3), 334-368.

Lau, D. C., & Murnighan, J. K. (1998). Demographic diversity and faultlines: The compositional dynamics of organizational groups. *Academy of Management Review, 23*(2), 325-340.

Lau, D. C., & Murnighan, J. K. (2005). Interactions within groups and subgroups: The effects of demographic faultlines. *Academy of Management Journal, 48*(4), 645-659.

Mannix, E., & Neale, M. A. (2005). What differences make a difference? The promise and reality of diverse teams in organizations. *Psychological Science in the Public Interest, 6*(2), 31-55.

Mao, A., Mason, W., Suri, S., & Watts, D. J. (2016). An experimental study of team size and performance on a complex task. *PloS One, 11*(4), e0153048.

Mazur, A., & Booth, A. (1998). Testosterone and dominance in men. *Behavioral and Brain Sciences, 21*(3), 353-363.

Mehta, P. H., Lawless DesJardins, N. M., van Vugt, M., & Josephs, R. A. (2017). Hormonal underpinnings of status conflict: Testosterone and cortisol are related to decisions and satisfaction in the hawk-dove game. *Hormones and Behavior*, *92*, 141-154.

Mehta, P. H., Wuehrmann, E. V., & Josephs, R. A. (2009). When are low testosterone levels advantageous? The moderating role of individual versus intergroup competition. *Hormones and Behavior, 56*(1), 158-162.

Meyer, B. & Glenz, A. (2013). Team faultline measures: A computational comparison and a new approach to multiple subgroups. *Organizational Research Methods,* 16, 393-424.

Murray, A. I. (1989). Top management group heterogeneity and firm performance. *Strategic Management Journal*, 10(S1), 125-141.

Phillips, K.W., & Apfelbaum, E.A. (2012). Delusions of homogeneity? Reinterpreting the effects of group diversity. *Research on Managing Groups and Teams, 15*, 185-207.

Phillips, K. W., Mannix, E. A., Neale, M. A., & Gruenfeld, D. H. (2004). Diverse groups and information sharing: The effects of congruent ties. *Journal of Experimental Social Psychology, 40*(4), 497-510.

Raftery, A. E. (1995). Bayesian model selection in social research. In P. V. Marsden (Ed.), *Sociological Methodology* (pp. 111-163). Cambridge, MA: Blackwell.

Ronay, R., Greenaway, K., Anicich, E. M., & Galinsky, A. D. (2012). The path to glory is paved with hierarchy: When hierarchical differentiation increases group effectiveness. *Psychological Science, 23*(6), 669-677.

Sommers, S. R. (2006). On racial diversity and group decision-making: Identifying multiple effects of racial composition on jury deliberations. *Journal of Personality and Social Psychology, 90*, 597–612.

Tajfel, H., & Turner, J. (1986). The social identity theory of inter group behavior. In S. Worchel & W. G. Austin (Eds), *Psychology of Intergroup Relations*. Chicago: Nelson.

Turner, J. C., Oakes, P. J., Haslam, S. A., & McGarty, C. (1994). Self and collective: Cognition and social context. *Personality and Social Psychology Bulletin, 20*(5), 454-463.

van Knippenberg, D., & Schippers, M. C. (2007). Work group diversity. *Annual Review of Psychology, 58*, 515-541.

Watson, W. E., Kumar, K., & Michaelsen, L. K. (1993). Cultural diversity's impact on interaction process and performance: Comparing homogeneous and diverse task groups. *Academy of Management Journal, 36*(3), 590-602.

Wheelan, S. A. (2009). Group size, group development, and group productivity. *Small Group Research, 40* (2), 247-262.

Wright, N. D., Bahrami, B., Johnson, E., Di Malta, G., Rees, G., Frith, C. D., & Dolan, R. J. (2012). Testosterone disrupts human collaboration by increasing egocentric choices. *Proceedings of the Royal Society of London B: Biological Sciences*, rspb20112523.

Zanutto, E. L., Bezrukova, K., & Jehn, K. A. (2011). Revisiting faultline conceptualization: Measuring faultline strength and distance. *Quality and Quantity, 45*(3), 701-714.