**SUPPLEMENTARY ONLINE APPENDICES**

**APPENDIX A**

**Study 1: Impact of Control Variables**

Following best practice recommendations (Becker et al., 2016), we re-ran our hypothesis tests with the control variables added to the models to assess their impact on our results and substantive conclusions. To assess the incremental validity of our national culture diversity variable, we controlled for functional diversity, which has been shown to impact team trust (magnitude). Respondents were asked to select the professional domain they had the most work experience in from a comprehensive list (e.g., ‘Accounting & Finance’, ‘Distribution, Warehousing & Logistics’). As with national culture diversity, functional background scores were subsequently aggregated to the team level using the same Blau index and formula. To assess the incremental validity of trust consensus on team performance, we followed Cole et al.’s (2011) best practice recommendations for testing dispersion-composition models and added team trust magnitude, a consensus-magnitude cross-product term, and quadratic terms for consensus and magnitude as control variables to our model in a stepwise manner. Doing so acknowledges that within-team dispersion and magnitude are systematically related (Lindell & Brandt, 2000), and accounts for the possibility of the two team trust properties affecting team performance interactively (e.g., De Jong & Dirks, 2012) rather than independently. Team trust magnitude was operationalized using the same data as our trust consensus measure, but applying mean-level aggregation instead. The mean rWG, ICC(1), and ICC(2) for trust magnitude were .91, .18, and .44, respectively. Our analysis showed a sizable correlation between trust magnitude and consensus (*r* = .45, see Table 1 in the paper), confirming the importance of accounting for their non-independence.

The results of our Hypothesis 1 test with control variables (Table 2, model 2, see paper) indicate the relationship between national culture diversity and trust consensus holds after controlling for functional diversity, with a point estimate (*B* = -.32) and CI (-.57, -.08) for national culture diversity being relatively unaffected by the addition of functional diversity. Similarly, the results in Table 3 (see paper) show the trust consensus-team performance relationship (Hypothesis 3) holds after controlling for trust magnitude (model 2), the consensus-magnitude interaction (model 3), and the curvilinear effects of consensus and magnitude (model 4), showing only modest changes in its point estimate (*B* = .48) and CI (.04, .90).

**Study 1: Impact of Team Tenure**

Our conceptualization of trust consensus as an emergent state presumes it has emerged and sufficiently endured over time and hence, can be sensed and experienced by team members (Waller et al., 2016). While it is reasonable to assume this for the subset of teams that had existed for a relatively long period (e.g., a year or more), this may not be evident for newly formed teams (e.g., a few months or less). Hence, we examined whether teams in our sample differed systematically in trust consensus as a function of their tenure. Team tenure was operationalized by members indicating their tenure in the current team (in years and months), and mean aggregating these scores to the team level using an additive model (Chan, 1998). We then took the model previously used for testing Hypothesis 1, and added both a main effect term for team tenure and a cross-product term between national culture diversity and team tenure in a stepwise manner. We mean-centered the lower-order terms prior to creating the higher-order term to facilitate interpretation of the interaction effect (Aiken & West, 1991).

The CI and point estimate for the team tenure main effect term (see Table A1, model 3), shows no meaningful direct relationship between team tenure and trust consensus, with parameter estimates including zero, and the most plausible estimate being -.08. Similarly, the CI for the national culture diversity-team tenure cross-product term includes zero and shows a considerable range of plausible estimates, with a point estimate of .19. These results suggest no meaningful interaction effect between national culture diversity and team tenure in predicting team trust consensus. Taken together, these results do not support the notion that team tenure systematically affects team trust consensus, suggesting the way in which team trust consensus emerged and endured in newly formed teams was similar to that of teams that had existed for a longer period of time.

|  |
| --- |
| Table A1 |
| Study 1: Regression Results for Team Tenure and Trust Consensus |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | Dependent variable: Team trust consensus |
| Independent variables |  | Model 1 |  | Model 2 |  | Model 3 |
|  |  | B | *SE* | CIL | CIU | *p* |  | B | *SE* | CIL | CIU | *p* |  | B | *SE* | CIL | CIU | *p* |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| National culture diversity (NCD) | -0.33 | 0.12 | -0.57 | -0.10 | 0.006 |  | -0.31 | 0.12 | -0.56 | -0.07 | 0.012 |  | -0.31 | 0.12 | -0.55 | -0.06 | 0.014 |
| Functional diversity |  |  |  |  |  |  |  | -0.07 | 0.12 | -0.30 | 0.16 | 0.564 |  | -0.07 | 0.12 | -0.31 | 0.16 | 0.526 |
| Team tenure (TT) |  |  |  |  |  |  |  | -0.08 | 0.09 | -0.27 | 0.10 | 0.359 |  | -0.08 | 0.09 | -0.26 | 0.10 | 0.386 |
| NCD \* TT |  |  |  |  |  |  |  |  |  |  |  |  |  | 0.19 | 0.25 | -0.31 | 0.70 | 0.448 |
| *R2* |  | 0.06 |  |  |  |  |  | 0.07 |  |  |  |  |  | 0.08 |  |  |  |  |
| *R2, p* |  |  |  |  |  |  |  | 0.01 | 0.568 |  |  |  |  | 0.00 | 0.448 |  |  |  |
| *F* |  | 7.75 |  |  |  |  |  | 2.94 |  |  |  |  |  | 2.34 |  |  |  |  |
| *N* = 120 teams. B = unstandardized regression coefficient; *SE* = standard error; CIL= 95% confidence interval lower limit; CIU = 95% confidence interval upper limit; *p* = exact p-value. |

**APPENDIX B**

**Study 2: Impact of Control Variables**

As with Study 1, we re-ran our hypothesis tests with the control variables added to the models to assess their impact on our results and substantive conclusions. Consistent with Study 1, we again controlled for functional diversity (operationalized as per Study 1) in testing the antecedent part of our model, and for trust magnitude, the consensus-magnitude interaction, and each of their curvilinear terms in testing the consequences part of our model (Cole et al., 2011). Team trust magnitude was operationalized using the same Time 3 data as our trust consensus measure, but applying mean-level aggregation instead. The mean rWG, ICC(1), and ICC(2) were .92, .18, and .56, and there was a sizable correlation between trust magnitude and consensus (r = .41, Table 4 in the paper). In addition, we also controlled for trust consensus at Time 2, the overall frequency of communication, leadership network decentralization, total amount of work experience, and prior familiarity. First, since emergence processes unfold along particular trajectories, progressing towards either greater consensus or greater dissensus over time (Kozlowski & Klein, 2000) early trust consensus should be predictive of later trust consensus. We therefore controlled for trust consensus at Time 2, operationalized as per our Time 3 measure, but using Time 2 data instead. Second, we controlled for overall communication frequency to exclude this as an alternative explanation for our Hypothesis 3 results. This was operationalized by taking the unweighted sum of respondents’ frequency ratings across four communication modes (the three virtual modes and face-to-face communication) and mean aggregating those to the team level. The single-item face-to-face communication measure used the same item stem and response anchors as the items used to operationalize team virtuality. Third, acknowledging that network density only captures part of the leadership structure in teams (D’Innocenzo et al., 2016), we controlled for network decentralization, operationalized using the same data as network density but applying a decentralization formula instead (Lemoine, Koseoglu, Ghahremani, & Blum, 2020). Fourth, as our sample included full- and part-time students who differed systematically in their work experience, we controlled for work experience, operationalized by mean-aggregating respondents' amount of work experience at Time 1 to the team level. Finally, although the teams were newly formed, we cannot exclude the possibility that some team members already knew each other prior to joining the team. We therefore controlled for prior familiarity by asking respondents at Time 1 the extent to which they already knew each teammate on a scale of 0 (“not at all”) to 2 (“very well”) and mean-aggregating these dyadic scores to the team level.

The results of the test of Hypothesis 1 with control variables (Table 5, model 3, see paper) show the range of plausible values for the negative effect of national culture diversity now includes zero (CI = -.46, .12) after adding in these controls. Instead of uncritically assuming the results of this model are more correct (Becker et al., 2016) and inferring a lack of support for Hypothesis 1, we acknowledge that both the point estimate and the CI require thoughtful interpretation (Amrhein, Greenland, & McShane, 2019). As such, we note that the most plausible estimate – the point estimate – is still negative (*B* = -.17) and directionally consistent with the results of the model without control variables (model 2). Furthermore, the vast majority of the values within the CI’s range (79%) are consistently negative, with the small portion of the range exceeding zero being located near the CI’s upper bound, making them disproportionally less plausible. As such, while not conclusive, the evidence still suggests it is more likely to be a negative effect of national culture diversity than not.

Regarding the trust consensus-team performance relationship (Hypothesis 2), the results of our re-analysis (Table 6, see paper) show that adding trust magnitude (model 2), the consensus-magnitude cross-product term (model 3), and the two curvilinear terms (model 4) did not alter our substantive conclusion regarding this relationship, with the point estimate (*B* = 3.77) still being directionally consistent with the results without control variables and the CI still not including zero (CI = .78, 6.76). As for our hypothesized interaction effects of national culture diversity with team virtuality (Hypothesis 3) and collective leadership density (Hypothesis 4), the results in Table 5 (model 3, see paper) show that these effects still hold after adding the controls, as the point estimates for the interactions are still directionally consistent (virtuality: *B* = .26; collective leadership: *B* = 6.21), their range of plausible values still excludes zero (virtuality: CI = .11, .42; collective leadership: CI = .84, 11.58), and both effect size magnitudes are above-average in magnitude (virtuality: *sr* = .31; collective leadership: *sr* = .21).

**Study 2: Impact of Virtuality Measure**

While the weighting of different communication technologies is desirable *conceptually* to capture asynchronicity and information value dimensions of team virtuality (Kirkman & Mathieu, 2005), it can have undesirable *statistical* implications (Evans, 1991). In particular, as the anonymous methods reviewer on our manuscript pointed out, the specific weighting values are arbitrary, such that different weightings might produce different results[[1]](#endnote-1). To assess the impact of the weightings and thus robustness of the results, we re-ran the model using an unweighted virtuality measure. The measure was identical to our original operationalization, except we did not differentially weight the different communication technologies. This re-analysis, reported in Table B1 model 2, shows directions of effect size estimates that are consistent with the pattern of interaction we hypothesized for national culture diversity and team virtuality (Hypothesis 3): negative for national culture diversity (*B* = -.36) and positive for team virtuality (*B* = .03) and the cross-product term (*B* = .43). Furthermore, the range of plausible values for the cross-product term still excludes zero (CI = .15, .72), and the effect size remains above average in magnitude (*sr* = .28). Adding control variables to the model (Table B1, model 3) produced nearly identical results (*B* = .45; CI = .15, .75; *sr* = .28). In sum, our re-analysis using the unweighted measure effectively replicates our original results, thereby attesting to the robustness of these results.

|  |
| --- |
| Table B1 |
| Study 2: Regression Results for Antecedent Factors Impacting Trust Consensus (Unweighted Virtuality Measure) |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | Dependent variable: Team trust consensus, Time 3 |
| Independent variables |  | Model 1 |  | Model 2 |  | Model 3 |
|  |  | B | *SE* | CIL | CIU | *p* |  | B | *SE* | CIL | CIU | *p* | *sr* |  | B | *SE* | CIL | CIU | *p* | *sr* |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| National culture diversity (NCD) |  | -0.23 | 0.13 | -0.50 | 0.04 | 0.090 |  | -0.36 | 0.13 | -0.62 | -0.10 | 0.008 |  |  | -0.22 | 0.15 | -0.53 | 0.09 | 0.155 |  |
| Team virtuality, unweighted (TVu) |  | 0.03 | 0.04 | -0.04 | 0.11 | 0.355 |  | 0.03 | 0.04 | -0.04 | 0.10 | 0.417 |  |  | 0.05 | 0.06 | -0.07 | 0.18 | 0.390 |  |
| Collective leadership density (CLD) |  | 1.34 | 0.63 | 0.09 | 2.59 | 0.036 |  | 0.83 | 0.61 | -0.38 | 2.04 | 0.175 |  |  | 0.66 | 0.64 | -0.62 | 1.94 | 0.306 |  |
| NCD\*TVu |  |  |  |  |  |  |  | 0.43 | 0.14 | 0.15 | 0.72 | 0.003 | 0.28 |  | 0.45 | 0.15 | 0.15 | 0.75 | 0.004 | 0.28 |
| NCD\*CLD |  |  |  |  |  |  |  | 6.27 | 2.67 | 0.97 | 11.57 | 0.021 | 0.22 |  | 7.13 | 2.78 | 1.60 | 12.67 | 0.012 | 0.24 |
| Team trust consensus, Time 2 |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 0.18 | 0.11 | -0.03 | 0.40 | 0.090 |  |
| Functional diversity |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 0.05 | 0.38 | -0.70 | 0.80 | 0.899 |  |
| Overall communication frequency |  |  |  |  |  |  |  |  |  |  |  |  |  |  | -0.02 | 0.05 | -0.13 | 0.08 | 0.636 |  |
| Collective leadership decentralization |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 0.00 | 0.42 | -0.83 | 0.83 | 0.996 |  |
| Prior familiarity |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 0.05 | 0.27 | -0.49 | 0.59 | 0.858 |  |
| Work experience |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 0.02 | 0.02 | -0.01 | 0.05 | 0.173 |  |
| *R2* |  | 0.11 |  |  |  |  |  | 0.22 |  |  |  |  |  |  | 0.27 |  |  |  |  |  |
| *R2, p* |  |  |  |  |  |  |  | 0.12 | 0.002 |  |  |  |  |  | 0.05 | 0.490 |  |  |  |  |
| *F* |  | 3.57 |  |  |  |  |  | 5.13 |  |  |  |  |  |  | 2.81 |  |  |  |  |  |
| *N* = 95 teams. B = unstandardized regression coefficient; *SE* = standard error; CIL= 95% confidence interval lower limit; CIU = 95% confidence interval upper limit; *p* = exact p-value; *sr* = semi-partial correlation. |

**REFERENCES**

Aiken, L. S., & West, S. G. 1991. *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Amrhein, V., Greenland, S., & McShane, B. 2019. Scientists rise up against statistical significance. *Nature*, 567: 305-307.

Lemoine, G. J., Koseoglu, G., Ghahremani, H., & Blum, T. C. 2020. Importance-weighted density: A shared leadership illustration of the case for moving beyond density and decentralization in particularistic resource networks. *Organizational Research Methods*, 23: 432-456.

1. We would like to thank the anonymous methods reviewer for raising this excellent point. [↑](#endnote-ref-1)