Running head: Hierarchy and Team Effectiveness Meta-Analysis

Why and When Hierarchy Impacts Team Effectiveness:
A Meta-Analytic Integration

Lindred L. Greer*, Bart A. de Jong, Maartje E. Schouten, and Jennifer Dannals

Journal of Applied Psychology, in press

* To whom correspondence should be addressed. Address: Stanford Graduate School of Business, Knight Management Center, 655 Knight Way, Stanford, CA, 94305. E-mail: lgreer@stanford.edu.
**We would like to think our reviewers as well as Drew Carton, Katy DeCelles, Frank Flynn, Nir Halevy, John Hollenbeck, Russ Johnson, Lara Tiedens, and Lisanne Van Bunderen for their constructive comments in developing this article.
*** Financial support was granted to the first author by the Netherlands Science Foundation (451-11-034) and to the second author by the Netherlands Science Foundation (016-145-243).

©American Psychological Association, 2017. This paper is not the copy of record and may not exactly replicate the authoritative document published in the APA journal. Please do not copy or cite without author's permission. The final article is available, upon publication, at (awaiting DOI).
Abstract

Hierarchy has the potential to both benefit and harm team effectiveness. In this paper, we meta-analytically investigate different explanations for why and when hierarchy helps or hurts team effectiveness, drawing on results from 54 prior studies \((N = 13,914\) teams). Our findings show that, on net, hierarchy negatively impacts team effectiveness (performance: \(\rho = -.08\); viability: \(\rho = -.11\)), and that this effect is mediated by increased conflict-enabling states. Additionally, we show that the negative relationship between hierarchy and team performance is exacerbated by aspects of the team structure (i.e., membership instability, skill differentiation) and the hierarchy itself (i.e., mutability), which make hierarchical teams prone to conflict. The predictions regarding the positive effect of hierarchy on team performance as mediated by coordination-enabling processes, and the moderating roles of several aspects of team tasks (i.e., interdependence, complexity) and the hierarchy (i.e., form) were not supported, with the exception that task ambiguity enhanced the positive effects of hierarchy. Given that our findings largely support dysfunctional views on hierarchy, future research is needed to understand when and why hierarchy may be more likely to live up to its purported functional benefits.

Keywords: Hierarchy, Teams, Meta-Analysis, Power, Status, Conflict
Why and When Hierarchy Impacts Team Effectiveness:

A Meta-Analytic Integration

Hierarchy, defined as vertical differences between members in their possession of socially valued resources (Harrison & Klein, 2007; Hays & Bendersky, 2015), is a fundamental concept in the study of groups and teams (for reviews, see Anderson & Brown, 2010; Halevy, Chou, & Galinsky, 2011; Keltner, Van Kleef, Chen, & Kraus, 2008; Magee & Galinsky, 2010). Past theory and research offers divergent perspectives on how and why hierarchy impacts team effectiveness. The functionalist perspective proposes a positive pathway from hierarchy to team effectiveness via improved coordination-enabling processes (e.g., Halevy et al., 2011; Keltner et al., 2008; Van der Heijden, Potters, & Sefton, 2009), while the conflict perspective proposes a negative pathway via increased conflict-enabling states (e.g., Bloom, 1999; Greer & Van Kleef, 2010; Greer, Van Bunderen, & Yu, 2017; Wolfe & McGinn, 2005). To understand these potentially off-setting processes and identify when one or the other is more likely to take priority, a growing line of work has started to identify different contingency factors that explain when hierarchy is most likely to help or hurt team effectiveness (e.g., Anderson & Willer, 2014; Bunderson, Van der Vegt, Cantimur, & Rink, 2016; Hays & Bendersky, 2015; Tarakci, Greer, & Groenen, 2016). These contingency factors include task characteristics (e.g., interdependence; Ronay, Greenaway, Anicich, & Galinsky, 2012), team structure (e.g., skill differentiation; Young-Hyman, 2017), and team hierarchy (e.g., hierarchy form; Bunderson et al., 2016).

While scholarly work in this area is burgeoning, the field currently lacks a cumulative investigation of the different, and potentially off-setting, explanations for why hierarchy impacts team effectiveness. Studies have thus far progressed in a disconnected and piecemeal fashion, with studies focusing on one pathway at a time, to the exclusion of others (e.g., Bunderson et al.,
This is problematic, as singling out a specific mechanism risks overstating the importance of that mechanism (De Jong & Elfring, 2010), prevents scholars from gauging the relative adequacy of different mediating mechanisms (Colquitt et al., 2013), and does not allow scholars to examine the extent to which these mechanisms complement or offset each other. Similarly, the field lacks a cumulative investigation of contingency factors explaining the conditions under which hierarchy helps or hurts team effectiveness, as research has also only tended to examine a single contingency factor at a time (e.g., Ronay et al., 2012). Given the multitude of contingency factors proposed so far, this lack of integrative testing may contribute to the proliferation of unnecessary theoretical complexity (Banks, McCauley, Gardner, & Guler, 2016), by preventing scholars from gauging the relative predictive power of any specific moderator (Ng, 2017).

Therefore, while the rapidly expanding work in this area is encouraging, the time has come to integrate and empirically compare the multiple proposed mediating mechanisms as well as integrate and empirically compare the multiple contingency factors (Bergh et al., 2014; Chan & Arvey, 2012).

The aim of this study is two-fold. First, we seek to integrate the two key pathways explaining why hierarchy has been proposed to impact team effectiveness into a single theoretical and empirical model to better understand how these pathways may co-exist and offset each other. Second, we integratively examine and compare a comprehensive set of contingency factors proposed by scholars to explain when the positive versus negative effects of hierarchy on team effectiveness will be more likely to take priority (e.g., Anderson & Brown, 2010; Anderson & Willer, 2014; Bunderson et al., 2016; Hays & Bendersky, 2015; Halevy et al., 2011). We consider these issues of why and when independently, and empirically test each of them via a meta-analysis, which integrates findings from 54 primary studies on hierarchy and
team effectiveness (involving 86 effect sizes, and 13,914 teams) conducted over the past six decades. In doing so, we aspire to provide the field with improved clarity, consistency, and predictive ability in construct and theory development, and help the field move away from theory proliferation absent explanatory power (Banks et al., 2016; Leavitt, Mitchell, & Peterson, 2010; LePine & Wilcox-King, 2010; Schlaegel & Koenig, 2014). Such integration can help provide insight into which perspectives on hierarchy may need qualification and give suggestions for how other perspectives may need changed or qualified.

**Why Hierarchy Impacts Team Effectiveness**

We seek here to identify and integrate the key explanations for why, as well as for when, hierarchy impacts team effectiveness. Consistent with the teams literature (Hackman, 1987; Kozlowski & Ilgen, 2006), team effectiveness is represented in this paper by two key indicators – team performance, or the degree to which a team accomplishes its goals or mission, as reflected by task-indicators such as team output quality, quantity, and efficiency (Bell, 2007; Devine & Phillips, 2001), and team viability, or the ability of the team to continue in the future based on member willingness to remain with the team, as reflected by socio-affective indicators, such as team satisfaction, commitment, and turnover intentions (Balkundi & Harrison, 2006).

In explaining the hierarchy-effectiveness relationship, the functionalist perspective proposes a positive pathway between hierarchy and team effectiveness via improved coordination-enabling processes (e.g., Anderson, Srivastava, Beer, Spataro, & Chatman, 2006; Keltner et al., 2008; Van Vugt et al., 2008), which are defined as the behaviors and strategies used to integrate and align individual member actions, knowledge, and objectives towards the attainment of common goals (Arrow, McGrath, & Berdahl, 2000; Rico, Sanchez-Manaznares, Gil, & Gibson, 2008; Zalesny, Salas, & Prince, 1995). The functionalist line of thought stems
from research in social psychology, which suggests that people have an unconscious preference for hierarchy (Tiedens, & Fragale, 2003) because of the comfort it gives in coordinating team interactions (Halevy et al., 2011). Scholars in this line of thought argue that hierarchy may benefit team effectiveness by facilitating and coordinating member interactions (e.g., De Kwaadsteniet, Van Dijk, Wit, De Cremer, & de Rooij, 2007; Galinsky, Chou, Halevy, & Van Kleef, 2012; Woolley, Gerbassi, Chabris, Kosslyn, & Hackman, 2008). This can include guiding resource allocation (Keltner et al., 2008), allowing members a means of knowing their place within the team (Anderson & Brown, 2010), and clarifying expectations about norms, roles, and expected behaviors based on members’ placement in the hierarchy (e.g., Biggart & Hamilton, 1984; Dornbusch & Scott, 1975). By providing a structure to guide roles and interactions, hierarchy may enhance coordination-enabling processes and thereby improve team effectiveness. For example, Woolley and her colleagues (2008) found in a laboratory experiment of student teams, that when teams had a clear expertise hierarchy, they were better able to integrate information and solve a hypothetical terrorist plot. Such coordination-enabling processes in turn benefit team effectiveness by facilitating expedient and effective combining of individual inputs to a team outcome (e.g., Cohen & Bailey, 1997; Gladstein, 1984). Indeed, meta-analytic evidence has shown that coordination-enabling processes have a positive impact on both team performance and viability (LePine, Piccolo, Jackson, Mathieu, & Saul, 2008). Taken together, these theoretical arguments and empirical evidence suggest that hierarchy can offer teams a positive pathway to team effectiveness through improved coordination-enabling processes.

In contrast, the conflict perspective proposes a negative pathway from hierarchy to team effectiveness via heightened conflict-enabling states (e.g., Bunderson & Reagans, 2011; DeBrock, Hendricks, & Koenker, 2004; Greer et al., 2017; Taracki et al., 2016), which are
defined as emergent team states characterized by the potential for perceived incompatibilities or differences between members (De Dreu & Gelfand, 2008; De Wit et al., 2012; Jehn, 1995). According to this perspective, hierarchy may motivate individuals to climb up the ranks (Magee & Galinsky, 2008), and may lead members at different ranks to have opposing interests and perspectives (Van Bunderen, Greer, & Van Knippenberg, 2017), which can lead to increased intragroup conflicts (over rank: Anderson & Brown, 2010; Greer & Dannals, 2017; Hays & Bendersky, 2015; and generally: Bunderson et al., 2016). Such conflicts have been shown to harm team outcomes (e.g., conflicts over rank: Bendersky & Hays, 2012; Ronay et al., 2012; conflicts more generally: for meta-analyses, see De Dreu & Weingart, 2003; De Wit et al., 2012), distracting members from task accomplishment and harming interpersonal relationships. In support of this perspective, for example, Greer and Van Kleef (2010) found that hierarchy harmed team performance in high-power teams because hierarchy provoked contests and conflicts over the rank order in the team. Indeed, a number of empirical studies have indeed now shown that hierarchy can create a tense intra-team environment, in which teams have high levels of conflict states and lowered team outcomes (e.g., Edmondson, 2002; Mannix, 1993; Torrance, 1955; Tost, Gino, & Larrick, 2013; Van der Vegt, De Jong, Bunderson, & Molleman, 2010; Wolfe & McGinn, 2005). Taken together, this line of work suggests that hierarchy can harm team effectiveness via conflict-enabling states.

Given the co-existence of these two off-setting pathways in the literature, we propose:

**Hypothesis 1:** Coordination-enabling processes mediate a positive relationship between hierarchy and (a) team performance and (b) team viability, even after controlling for the mediating effect of conflict-enabling states.

**Hypothesis 2:** Conflict-enabling states mediate a negative relationship between hierarchy
and (a) team performance and (b) team viability, even after controlling for the mediating effect of coordination-enabling processes.

**When Hierarchy Impacts Team Performance**

Recognizing the need to reconcile the off-setting processes which have been proposed to explain why hierarchy affects team effectiveness, and to understand when one process will take priority over the other, scholars have argued for the need to adopt a ‘bounded account of hierarchy’ (e.g., Anderson & Willer, 2014; Bunderson et al., 2016), and have begun to identify contingency factors that could explain the conditions under which hierarchy is beneficial or detrimental for teams. For example, Halevy, Chou, Galinsky, and Murnighan (2012) suggest that hierarchy benefits performance in teams with high coordination demands (i.e., high task interdependence), and Bunderson et al. (2016) show that the conflictual downsides of hierarchy are worse when the hierarchical form is based more on inequity or centralization than acyclical patterns of dyadic influence. Given that the majority of contingency findings have been proposed for the relationship between hierarchy and performance (and not viability), we focus our theorizing here on team performance, but encourage future research to examine contingencies of the hierarchy–team viability relationship as well.

In integrating specific contingency factors, we purposefully build on pre-existing contingency frameworks that have been proposed in the literature (e.g., Anderson & Brown, 2010; Bunderson et al., 2016; Greer et al., 2017; Hays & Bendersky, 2015; Halevy et al., 2011). To ensure that our contingency factors are comprehensive and meaningful as a set, we focus on categories of factors that are conceptually distinct in nature and functionally distinct in terms of which pathway between hierarchy and team performance they cause to dominate. With regard to conceptual distinctions, we draw on past team meta-analyses and differentiate between
contingency factors that pertain to aspects of team tasks (e.g., task interdependence, LePine et al., 2008), aspects of team structure (e.g., skill differentiation, De Jong, Dirks, & Gillespie, 2016), and aspects of the focal independent variable itself (e.g., form of diversity, Bell et al., 2011), in our case hierarchy. With regard to functional distinctions, we link our contingency factors to the two off-setting pathways identified in the literature and differentiate between factors that are predicted to give priority to the positive pathway between hierarchy and performance by increasing coordination demands (e.g., task complexity, Bunderson et al., 2016), and factors that are predicted to give priority to the negative pathway by increasing conflict susceptibilities (e.g., hierarchy mutability, Hays & Bendersky, 2015). Below, we describe each specific contingency factor, and our predictions for how it will affect the relationship between hierarchy and team performance.

**Team Task Contingencies**

Theory and research on hierarchy has led to the expectation that hierarchy will most benefit performance when teams work on tasks with high coordination demands (Anderson & Brown, 2010; Halevy et al., 2011). Not all teams require coordination to perform well, but for teams where aspects of the team’s task place high demands on members to integrate and align individual member actions, knowledge, and objectives towards the attainment of common goals (Arrow et al., 2000; Rico et al., 2008), we expect hierarchy to be the most useful. Hierarchy can help meet coordination demands by providing a heuristic to allocate resources in the team (Keltner et al., 2008), to integrate individual efforts as members will know who to defer to for what (Halevy et al., 2011), and to process information and provide a structure to organize information flow, ensuring no valuable knowledge gets lost. When teams in high need of coordination are able to coordinate individual contributions well, they are likely to perform better
(Steiner, 1972). Thus, as implied by the functionalist perspective, hierarchy should benefit teams when coordination demands are high. In contrast, in teams with low coordination demands, hierarchy may become superfluous and unlikely to positively impact team performance.

We focus here on three aspects of team tasks which increase the utility of such coordination-enabling processes and have been commonly investigated in the hierarchy literature: task interdependence, task ambiguity, and task complexity. Task interdependence is defined as the degree to which members need to rely on and connect with one another for task accomplishment (Shea & Guzzo, 1987; Saavedra, Earley, & Van Dyne, 1993). Task interdependence increases coordination demands on teams (Saavedra et al., 1993), as when members are closely bound together to accomplish team objectives, they have to communicate more frequently, make common decisions on team task components, and sync individual task efforts with each other. In such situations, the structure and mental-map provided by hierarchy can help coordinate these interactions, enabling higher team performance. Scholars in this area support this reasoning, as they have theorized and found that hierarchy is more positively related to performance when interdependence is high (e.g., Halevy et al., 2012; Ronay et al., 2012).

Task ambiguity is defined as the degree to which members lack clarity about their responsibilities and what is expected of them to accomplish a task (Kahn, Wolfe, Quinn, Snoek, & Rosenthal, 1964). Task ambiguity can increase the coordination demands on the team (Barton, 1980; Valentine & Edmondson, 2014). When task ambiguity is high, team members have high levels of uncertainty about how they should exert their work-related efforts, and how their efforts should align with others. In such situations, hierarchy can provide a heuristic to help guide and structure member activities and interactions. Given its relevance to the study of hierarchy and structure, task ambiguity has indeed been proposed as a moderator of the effects of
hierarchy on team performance (e.g., Pandey & Rainey, 2006).

Task complexity implies the presence of multiple potential pathways to achieve a goal, the presence of multiple desired outcomes, the presence of conflicting interdependence among the different paths to achieve goals, and the presence of probabilistic links between paths and outcomes (Campbell, 1988). Task complexity has been shown to increase the coordination demands on teams (e.g., Zhou, 2013). When task complexity is high, team members have a high information load, and anything which can simplify their responsibilities and guide their interactions can be useful to facilitate task accomplishment. As such, task complexity has indeed been shown to moderate the effects of hierarchy on performance (e.g., Bunderson et al., 2016).

Consistent with this theorizing and empirical evidence, we hypothesize:

**Hypothesis 3**: Task interdependence (a), task ambiguity (b), and task complexity (c) moderate the positive relationship between hierarchy and performance, such that the relationship becomes stronger as the levels of these task characteristics increase.

**Team Structure Contingencies**

In understanding how team structure may affect the relationship between hierarchy and team performance, we build on emerging trends in the hierarchy literature which suggest that hierarchy’s downsides are most likely when conflicts emerge in the team (Bunderson et al., 2016; Greer & Van Kleef, 2010). While not all teams are particularly prone to conflict, in teams whose structures leave them particularly vulnerable to conflict, we expect to see the most negative effect of hierarchy on team performance. The reasoning in this work derives from the conflict perspective on hierarchy (Tarakci et al., 2016), which suggests that in teams whose structures are susceptible to conflicts, the rank differences created by hierarchy may be especially likely to trigger intense conflicts and team performance is likely to suffer. Namely,
when teams are already susceptible to conflict, hierarchies may become a bone of contention to activate underlying tensions, such as between different functional subgroups in highly skill-differentiated teams. In such situations, differences in ranks and resources may be particularly salient and resented, as they will be paired with underlying categorical in-group/out-group distinctions, which can cause members to engage in conflicts over ranks and resources. Such hierarchical conflicts have been shown to distract members, impair conflict resolution process, and harm team performance (Bendersky & Hays, 2012; Chun & Choi, 2014; Greer & Van Kleef, 2010). Therefore, teams with structures more susceptible to internal conflict-enabling states are more likely to have negative effects of hierarchy on team performance.

Besides hierarchy itself (what Hollenbeck, Beersma, & Schouten [2012] term ‘authority differentiation’), we focus on two other team structural characteristics that together comprise the big-three team structural characteristics, namely membership instability and skill differentiation. Membership instability is defined as teams with frequent membership changes (Hollenbeck et al., 2012). When teams have frequent membership changes, and members enter or leave or the team as a whole forms and dissolves quickly, the hierarchy has to be frequently renegotiated. This ongoing conflict can lead to rivalries and tensions among different members, and the presence of a hierarchy in such a volatile situation can explode existing tensions and fuel future rivalries. Such frequently changing teams may instead benefit from flatter structures, where less negotiation is required as members enter or leave the team. Membership instability has been shown to increase a team’s conflict susceptibilities (e.g., Tuckman, 1965), and has been suggested to have implications for the utility of hierarchical team structures (e.g., Valentine & Edmondson, 2014).

Skill differentiation is defined as the degree to which members have specialized
knowledge or functional capacities that make it difficult for members to be substituted and this is reflected in differences in members’ educational major, area of functional work experience, and/or expertise area (Hollenbeck et al., 2012). Skill differentiation has been shown to increase a team’s conflict susceptibilities (Bunderson & Sutcliffe, 2002; Cronin & Weingart, 2007; Dougherty, 1992). When teams consist of members with different skill sets, underlying tensions can exist between different functional groups (Dougherty, 1992). Hierarchy can then trigger such tensions by creating inequities and resentments between these different subgroups. Skill differentiation has indeed been suggested in the literature on hierarchy to have implications for the utility and effectiveness of hierarchical structures (e.g., Gruenfeld & Tiedens, 2010; Young-Hyman, 2017). Based on these lines of theorizing and empirical findings, we predict:

**Hypothesis 4:** Membership instability (a) and skill differentiation (b) moderate the negative relationship between hierarchy and performance, such that the relationship becomes stronger as the levels of these structural characteristics increase.

**Team Hierarchy Contingencies**

An additional important set of contingency factors that may determine whether teams are helped or harmed by hierarchical differentiation are aspects of the hierarchy itself. We define hierarchy in this paper broadly as vertical differences in socially valued resources. However, hierarchies may vary widely in the basis and structure of these vertical differences, and some forms and bases of hierarchy are more likely to be contested than others. When hierarchies motivate and/or allow for more individual mobility up or down the team hierarchy, such settings increase individuals’ concerns with their self-interest (Maner & Mead, 2010; Wright, Taylor, & Moghaddam, 1990) and the likelihood of hierarchy impairing team performance via hierarchical struggles and conflicts (e.g., Greer & Van Kleef, 2010).
Therefore, we focus here on two critical dimensions of hierarchy which measure the degree to which hierarchies are more likely to be contested or not – hierarchy mutability and hierarchy form. Hierarchy mutability is defined as the degree to which the basis of a hierarchy (e.g., status or power) is open to be contested and has been shown to increase a team’s conflict susceptibilities (Hays & Bendersky, 2015). When hierarchies are themselves easily changed, such as informal or status-based hierarchies, members may see more possibility for upward movement, and lower ranked members may seek to climb up, while higher ranked members seek to maintain rank, leading to rank-based conflicts between members (cf. Greer et al., 2017). Additionally, in such mutable hierarchies, more disagreements about hierarchical rank may exist, which can contribute to deteriorated individual contributions and performance-deterring intragroup conflicts (e.g., Greer, Caruso, & Jehn, 2011; Kilduff, Willer, & Anderson, 2012). Hierarchy mutability has indeed been able to explain when hierarchies are more likely to lead to performance-deterring conflicts (e.g., Hays & Bendersky, 2015).

Hierarchy form is the degree to which members within the team’s hierarchy are arrayed in an inequitable, centralized, or acyclical manner (Bunderson et al., 2016). Hierarchies whose forms evoke more perceived inequities and politics – i.e., steep and/or centralized hierarchies, have been suggested to evoke more conflict in hierarchical teams (e.g., Eisenhardt & Bourgeois, 1988; Harrison & Klein, 2007) than hierarchies whose form is more legitimized through accepted downwards flowing dyadic influence patterns (what Bunderson et al. [2016] term ‘acyclical’ hierarchies). When hierarchies are experienced as more illegitimate and political, such as steep or highly centralized hierarchies, they are more likely to promote negative team conflicts and hurt team performance. Hierarchical form has indeed recently been found to have implications for the conflict-susceptibilities of hierarchies (e.g., Bunderson, et al., 2016). Based
on this work, we propose:

**Hypothesis 5:** Hierarchy mutability (a) and hierarchy form (b) moderate the negative relationship between hierarchy and performance, such that the relationship becomes stronger as mutability increases and hierarchy is based on inequality (as opposed to acyclic).

**Method**

**Literature Search**

In order to identify all relevant studies to include in our meta-analytic test of the hypotheses presented here, we performed a combination of search strategies. First, we performed searches in several online databases (e.g., ABI/Inform, Google Scholar, PsycINFO, Web of Science) to find any articles that included *team* or *group* in conjunction with the keywords *hierarchy, hierarchical differentiation, authority differentiation, power, status, influence, leadership, expertise, inequality,* and *dispersion*. We also performed backward and forward searches on key theoretical articles on the topic (e.g., Halevy et al., 2011; Magee & Galinsky, 2008) to ensure we did not overlook any relevant studies. Complementing these electronic search strategies, we manually examined the table of contents from top-tier journals in psychology, management, and economics (e.g., *Academy of Management Journal, Administrative Science Quarterly, Applied Economics, Journal of Applied Psychology, Journal of Management, Journal of Organizational Behavior, Journal of Personality and Social Psychology, Organization Science,* and *Strategic Management Journal*) published in the past five years. Lastly, to minimize publication bias, we scanned the programs of relevant conferences and sent emails to the members of the related associations via their newsletters (e.g., Academy of Management, the Society for Industrial and Organizational Psychology, Interdisciplinary Network for Groups
Research, and the Society for Personality and Social Psychology), searched for unpublished dissertations through ProQuest, and sent requests for relevant working papers to key authors in the field. Our literature search includes articles available up until July 2015.

**Inclusion Criteria and Initial Sample**

Our literature search returned a list of over 1,500 potentially includable articles. In order to be included, studies needed to contain enough information to allow for the computation of effect sizes and report empirical measures of hierarchy, and team mediators (conflict-enabling states, coordination-enabling processes) or outcomes (performance, viability), or both, at the team-level of analysis. We included papers that reported effect size estimates for hierarchy and any one potential dependent variable (mediators or outcomes) – papers did not need to include all possible dependent variables. In addition, the hierarchy and dependent variable measures needed to be consistent with our definitions of our core constructs in order to be included in our sample.

Given that we define hierarchy here as vertical differences between individuals within a team with respect to the possession of socially valued resources or assets (cf. Harrison & Klein, 2007; Zitek & Tiedens, 2012), we included hierarchy measures that were based on different resources which are socially valued in work teams. We draw on work on expectation states theory (e.g., Berger, Cohen, & Zelditch, 1972; Bunderson, 2003) to identify resources most proximally related to task performance to be the most socially valued resources in teams, and we therefore include hierarchy measures which reflect such task-related resources, including job titles, leadership authority, and expertise levels. While other resources that are not directly related to the task may also give rise to intragroup hierarchy, such as gender and race (e.g., Bunderson, 2003), we excluded these from our sample because they have not been as
commonly recognized in the literature as proxies of intragroup hierarchy. For example, gender and race are generally operationalized in terms of variety-based diversity rather than as separation- or disparity-based diversity, precluding them from capturing team hierarchy (Harrison & Klein, 2007). Furthermore, some attributes, especially pay levels, are often not publicly known within the team, and do not have the same salience or relevance that internal team bases of hierarchy, such as status or job rank would. Given these conceptual boundaries, valid indicators for hierarchy included (but were not limited to), negotiation BATNAs (the value of the best alternative members had to reaching a negotiated agreement in that situation), influence networks, manipulations of the presence of leaders or layered hierarchies or not, and status round-robin for hierarchy.

For conflict-enabling states and coordination-enabling processes, valid indicators included self-report, video-coded, network-measures, and archival measures of the levels of conflict-enabling states and coordination-enabling processes occurring within the team. For example, several studies used self-report scales of conflict-enabling states (e.g., Martins, Schilpzand, Kirkman, Ivanaj, & Ivanaj, 2013), and for coordination-enabling processes, several studies used self-report measures, including the ability of teams to align actions, to piece together information and provide each other regular feedback (e.g., Greer, Dannals, & Rao, 2016; Woolley et al., 2008). Additionally, consistent with our broad definition of conflict-enabling states and coordination-enabling processes, we not only considered direct measures of conflict and coordination as valid indicators, but also included measures of power struggles and psychological safety (for conflict-enabling states) and measures of knowledge integration and information exchange (for coordination-enabling processes). For team outcomes, valid indicators for performance included leader ratings of team performance quality, financial earnings, and
matches won, and for team viability valid indicators included self-report measures of team satisfaction and commitment (cf., Balkundi & Harrison, 2006). Applying these inclusion criteria yielded a usable sample of 54 relevant studies, involving 86 effect sizes, and 13,914 teams. Thirty-nine of these studies were published, seven were unpublished dissertations, and eight studies involved new, unpublished data. All included studies are listed in the references section, marked by an asterisk, and an overview of the data obtained from each study can be found in Appendix A. For an overview of excluded studies, and the reasons for their exclusion, please see our online supplements.

Coding Procedures

Since the majority of the studies were correlational, we used Pearson’s correlation coefficient ($r$) as our effect size metric. When other metrics – such as t-tests from ANOVAs – were reported in studies, they were converted into correlations following formulas proposed by Lipsey and Wilson (2001). When correlations were based on negative indicators of one of the two correlated variables (i.e., such that higher scores indicated lower values of that variable), we reverse-coded those correlations to ensure coding consistency across correlations. When studies reported both lagged and cross-sectional correlations with the same mediator or outcome, we only included the lagged correlations in order to be consistent with the causal directionality implied in our framework (cf., De Jong et al., 2016). When multiple correlations of hierarchy with the same dependent variable were available from the same study, they were combined using linear composites to ensure independence of observations (Schmidt & Hunter, 2015), or averaged when the information reported in primary studies was insufficient for creating linear composites.
When correlations were based on subjectively measured variables, we corrected the correlations for measurement error using different reliability statistics, depending on the nature of the measure (cf., Courtright et al., 2015; Hulsheger, Anderson, & Salgado, 2009; Wang, Waldman, & Zhang, 2014). Specifically, for single-rater measures (e.g., team leader’s subjective rating of the team’s performance), we used the Cronbach’s Alpha to correct study correlations. For multi-rater measures (e.g., team members’ subjective ratings of team conflict), we used ICC(2) to correct study correlations for rater-specific measurement error whenever they were reported, or used ICC(1)s and F-statistics to calculate ICC(2)s ourselves (cf., De Jong et al., 2016). If none of these aggregation statistics were reported, however, we used Cronbach’s Alpha to at least correct for item-specific measurement error in team-level mean scores. When reliabilities were missing for subjective measures all together, we imputed the sample-size weighted mean reliability of all available studies (cf., Steel, 2007). When hierarchy was experimentally manipulated or when hierarchy or team processes or outcomes were measured objectively, their reliability was set to 1. Whenever we used linear composites to combine multiple correlations from the same study, we used Mosier reliability composites to combine internal consistency coefficients (Schmidt & Hunter, 2015).

While extracting ‘low inference’ data from the studies (i.e., statistics) was relatively straightforward, more detailed coding schemes needed to be developed to code ‘high inference’ data on our moderators. Almost all of our moderating variables were coded on a continuous scale ranging from 1 to 5, with higher scores reflecting higher levels of the variable. The only exception was hierarchy form, which was categorically coded as (1) centralization, (2) steepness, or (3) acyclicity. Based on the coding procedures, all moderators (across all 52 studies that reported a hierarchy-performance correlation) were independently coded by both the first and the
third author. The analyses showed a mean Krippendorf’s Alpha of .94 across moderators ($SD = .03$; Hayes & Krippendorff, 2007), suggesting high levels of inter-coder agreement. Any discrepancies between the authors were resolved via discussion until consensus was reached.

**Team Task Contingencies.** We operationalize *task interdependence* as the extent to which team members were able to divide the team’s task or the extent to which the team’s task required interaction to complete it (Courtright, Thurgood, Stewart, & Pierotti, 2015; Horwitz & Horwitz, 2007). We coded task interdependence as ‘low’ (1) when team members could divide their work into sub-tasks and work on their own and their efforts would be pooled together for a team outcome (e.g., Mannix, 1993), as ‘moderate’ (3) when team members worked together closely for some but not all tasks – members at times would come together to collaborate and at other times would go off to accomplish individual sub-tasks (e.g., He & Huang, 2011), and as ‘high’ (5) when team members had to continuously work closely together, as each member’s contribution was reliant on other team members’ contributions, and team outcomes were the result of the collective effort of all team members (e.g., Bunderson, 2003). This is consistent with other meta-analyses that have examined task interdependence, such as Horwitz and Horwitz (2007), and Joshi and Roh (2009). By capturing task interdependence in a more continuous fashion, we were able to account for both established models of interdependence, such as pooled, sequential, reciprocal, and comprehensive, and the complex realities of tasks that influence the nature of interactions among team members in ongoing teams (LePine et al., 2008). To infer task interdependence, we drew on information about the collaborative nature of work as described in the description of the sample and team task. For example, baseball teams, where members could function loosely from one another, were coded as having low interdependence, while management teams, where members at times function independently in their own units and at
times interdependently on management decisions, were coded as having moderate
interdependence, and student lab teams, where members are highly dependent on one another to
accomplish the team task, were coded as having high interdependence. The studies included in
this meta-analysis reflected all levels of task interdependence.

Kahn et al. (1964) argued that task ambiguity arises when an individual lacks the
necessary information about the task to carry out his or her role. Specifically, according to Kahn
et al. (1964), three subtypes of task ambiguity can arise related to a) the scope of responsibilities
(i.e. what is expected of me?), b) the behavioural responsibilities (i.e. what activities will lead me
to accomplish what is expected of me?), and c) the hierarchy of responsibilities (i.e. what tasks
are prioritized over others?). To our knowledge, we are the first to examine task ambiguity in a
meta-analytic fashion. We therefore leveraged this definition by Kahn et al. (1964) to develop a
coding scheme for task ambiguity. We coded task ambiguity as ‘low’ (1) when members had
full, complete information about their task responsibilities within the team, including the scope,
nature, and prioritization of their task responsibilities and behaviors within the team, as
‘moderate’ (3) when members had some clarification and boundaries around what behaviors and
activities were expected of them individually on the team task, and task ambiguity was coded as
high (5) when members lacked clarity about task responsibilities and expectations, including the
scope, nature, and prioritization of member responsibilities within the team. To infer task
ambiguity, we drew on information about the goal, task, and performance expectations of the
teams studied as described in the methods sections of the papers we coded. Settings, such as
teams of healthcare specialists, where the task and associated responsibilities were clearly known
to members, were coded as having low task ambiguity (Mitchell et al., 2015). In contrast,
settings in which teams did not have much knowledge about how to accomplish the task and had
to figure it out on their own, such as startup teams, where teams had to figure out who would do what and what needed to be done to reach a good decision, were coded as having high task ambiguity (Foo, Sin, & Yiong, 2006). When samples were composed of a mixture of teams, for instance due to snowball sampling, samples were coded as having moderate task ambiguity (Van der Vegt et al., 2010). Included studies reflected all levels of task ambiguity.

In contrast to task ambiguity, in which team members lack information about their role and how to prioritize their responsibilities, task complexity implies that members do have information about the task, but that the information load is high. Namely, task complexity arises when there is the presence of multiple pathways to meet a goal, the presence of multiple desired outcomes, the presence of conflicting interdependence among paths to multiple outcomes, and the presence of probabilistic links among paths or outcomes (Campbel, 1988). In short, task accomplishment is quite detailed and intricate, and in need of considerable thought and exploration. In coding task complexity, we leveraged this definition to develop a coding scheme. We coded task complexity as ‘low’ (1) when the team’s task was simple and a limited number of task activities were clearly linked to outcomes, as ‘moderate’ (3) when there were a few different goals or paths to the goals, but the task was relatively straightforward, and task complexity was coded as ‘high’ (5) when numerous goals, and/or paths to the goals existed, and considerable thought was required to achieve desired outcomes. To infer task complexity, we relied on the descriptions of the team’s task. Tasks, such as regularly scheduled team meetings, were coded as low in complexity due to their regular occurrence (Sauer & Kauffeld, 2013). Tasks, such as those of a top management team, were coded as high in complexity due to the presence of numerous subgoals and strategies to task accomplishment (Boone & Hendriks, 2009). All levels of task complexity were present in the dataset, with a larger number of moderately low complex tasks.
**Team Structure Contingencies.** Membership instability is one of the two factors that make up a team’s temporal stability, in addition to team longevity (Hollenbeck et al., 2012). Our coding scheme for membership instability was based on Lee, Koopman, Hollenbeck, Wang, and Lanaj’s (2015) team descriptive index short form. Lee and her colleagues developed a 9-point Thurstone response scale with 5 descriptive anchors. We reversed and compressed the 9-point scale to a 5-point scale by retaining the anchors for each of the levels and narrowing the focus of the descriptors to only reflect membership instability. Membership instability was coded as ‘low’ (1) when teams experienced very little membership change or turnover resulting in team members that were very familiar with each other and a team that rarely, if ever, recruited new members, as ‘moderate’ (3) when teams had some level of membership change, but whereby there was a main, core set of team members, and as ‘high’ (5) when teams had high levels of membership change and members were almost total strangers to each other or did they spend enough time together to really get to know one another. To infer membership instability, we drew on information on the brevity of the team’s time together and the permeability of team boundaries. Teams at all levels of membership instability were included in this meta-analysis. Management teams (Haleblian & Finkel, 1993), manufacturing teams (Sauer & Kauffeld, 2013), healthcare teams (Thylefors, 2012), semester-long student project teams (Martins et al., 2013), and one-shot lab teams (Tost et al., 2013) are examples of teams coded from low to high at each level of membership instability, respectively.

In the same paper, Lee et al. (2015) developed a 9-point Thurstone measure for skill differentiation with five descriptive anchors. Skill differentiation describes the extent to which members have specialized knowledge or functional capacities (Hollenbeck et al., 2012). Similar to our procedure for membership instability, we compressed the 9-point format to a 5-point scale
by retaining the anchors for each of the levels. Skill differentiation was coded as low (1) when team members’ roles, responsibilities, and skillsets were overlapping and/or highly substitutable with one another or the tasks were so simple that even untrained people could perform almost all of the tasks, as ‘moderate’ (3) when members had a core set of skills, roles, and responsibilities that they all shared but each member still had some degree of unique skills and contribution to the task, and as ‘high’ (5) when the team had a clear lateral division of labor, based on differences in members’ unique roles, responsibilities, and areas of expertise. The extent of skill differentiation was relatively evenly distributed across studies in our sample. The following are examples of teams ranging from low to high, respectively: Student lab teams (e.g., Lemoine, Koseoglu, & Blum, 2015), negotiation teams with different role instructions (Pinkley, Neale, & Bennett, 1994), R&D teams (Ren, Gray, & Harrison, 2015), and top management teams (Hambrick, Humphrey, & Gupta, 2015). To infer skill differentiation, we drew on information about the degree to which different specializations were present in the team.

**Team Hierarchy Contingencies.** Hays and Bendersky (2015) recently found that the extent to which team members perceive that they can change their position in the hierarchy drives more competitive and conflict prone behaviors in teams. They coined the term *hierarchy mutability* to describe the extent to which the basis for the team’s hierarchy allowed for the opportunity for upward mobility. We coded a team’s hierarchy mutability as ‘low’ (1) when team members were unlikely to be able to change their position in the team hierarchy, such as in the teams of medical clinic nurses examined by Perry (2014), in which job positions are generally stable. Teams were coded as ‘moderate’ (3) when team members had some potential to change their position in the hierarchy, such as the production teams examined by Bunderson (2003) which allowed for some potential in the change of the hierarchical ranking of members.
Finally, teams were coded as ‘high’ (5) in hierarchy mutability when team members could easily change their position in the team hierarchy, such as the start-up teams examined by Foo et al. (2006) which operated in a dynamic and complex environment in which team members were trying to find their roles. To infer hierarchy mutability, we drew on information on the entrenchment and type of hierarchy described in the study. While teams were identified at all levels of mutability, the distribution in our sample was clearly bimodal with the majority of teams scoring either toward the lower (1) and (2) or the higher (5) hierarchy mutability values.

Bunderson et al. (2016) argue that the way the hierarchy is structured matters for its outcomes. Following their reasoning, we coded *hierarchy form* in four different categories. Following Bunderson and colleagues, we distinguished between (1) centralization, (2) steepness, (3) acyclicity, and (4) mixed. The centralization approach conceptualizes hierarchy as the concentration of hierarchy characteristics (e.g., power, status) in one team member or a small group of team members. Centralization is captured through measures of concentration, such as the coefficient of variance or the Gini coefficient. Examples of scholars studying this type of hierarchy form are He and Huang’s (2011) measure of board membership inequality and Rulke and Galaskiewicz’ (2000) centralization of the teams’ network structures. The steepness approach conceptualizes hierarchy as the aggregate difference across members on hierarchy characteristics. Steepness is captured through measures of standard deviation and Euclidian distance. Examples of this type of hierarchy form are the standard deviation of power levels that Greer and Van Kleef (2010) used and the manipulation of the leadership role that Becker and Baloff (1969) used to study the effectiveness of different organizational structures. The most recently established form of hierarchy, acyclicity, is conceptualized as “cascading relations of dyadic influence (p. 1270)”. Acyclicity is operationalized as a network of directed influence.
relations, for instance, by calculating Krackhardt’s (1994) hierarchy index. Given the recent developments with regard to this form of hierarchy, we were able to uncover only two studies that operationalized hierarchy as acyclicity in terms of task-relevant resources, as we focused on here, and looked at group-level processes and outcomes (Bunderson et al., 2016; Cantimur, Rink, & Van der Vegt, 2015). Given the categorical nature of our coding scheme, we coded hierarchy form as ‘mixed’ when studies reported multiple forms of hierarchy.

Meta-Analytic Approach

We performed our meta-analysis using Version 3 of the Comprehensive Meta-Analysis software (Borenstein, Hedges, Higgins, & Rothstein, 2014), which relies on the Hedges and Olkin (1985) approach. Consistent with best practice recommendations (Geyskens, Kirshnan, Steenkamp, & Cunha, 2009), we used a random effects model, which assumes that variability in correlations across studies is caused by both sampling error and true effect size variability in the population (Schmidt & Hunter, 2015).

There is growing scholarly recognition that, rather than interpreting effect size magnitudes in absolute terms using universal standards (e.g., Cohen, 1988), effect size magnitudes should be interpreted in reference those found in prior research in related areas (Bosco, Aguinis, Singh, Field, & Pierce, 2015; Paterson, Harms, Steel, & Crede, 2016). Consistent with this recommendation, we relied on the average effect size estimate across team-level meta-analyses ($\rho = .26$) reported by Paterson et al. (2016) as our reference for interpreting our main effects, and the average effect size estimate across team-level meta-analyses (semi-partial correlations: $sr = .20$) reported by O'Boyle, Banks, Walter, Carter, and Weisenberger (2015) for interpreting our moderator effects. In absence of pre-existing benchmarks for mediation effect sizes, we constructed a benchmark ourselves by identifying seven prior team-
level meta-analyses that tested mediation effects (e.g., Knight & Eisenkraft, 2015), extracting the reported indirect effect size estimates for each meta-analysis (see online supplements), and then using the reported harmonic means to calculate the sample-size weighted average across those meta-analyses (marked by double asterisks in the reference section). This resulted in a mediated effect size benchmark of $ab = .09$. Accordingly, we interpret our effect sizes in reference to the above effect size estimates, and qualify them as ‘below-average’, ‘average’ and ‘above-average’ (cf., De Jong et al., 2016).

Besides interpreting the magnitude of our effect size estimates, we also consider the precision with which these estimates capture the true effect in the population by constructing 95% confidence intervals (CIs) around these estimates. A 95% CI represents an estimation of the range of plausible values of the true (unknown) population parameter, and indicates that we can be 95% confident that the interval’s lower and upper bounds contain that population parameter (Cumming, 2012; Ellis, 2010). The narrower the CI, the more precise the point estimate is in capturing the true population parameter. CIs can also be interpreted as indicators of statistical significance (Cummings, 2012). Specifically, a 95% CI that does not include zero may be interpreted as a statistically significant result at the level of $p < .05$.

As meta-analytic findings can be severely compromised by publication bias, we first assessed the presence of such bias in our data using multiple techniques, namely the Begg and Mazumdar’s rank correlation test, the Egger’s intercept test, and Duval and Tweedie’s (2000) trim-and-fill method (for more details of each test, see Kepes, Banks, McDaniel, & Whetzel, 2012). The Begg and Mazumdar’s rank correlation test was non-significant ($\tau = -.16, p = .09$), as was Egger’s regression intercept test ($\beta_0 = -.19, p = .59$). Furthermore, both the Duval and Tweedie’s Trim and Fill method and associated funnel plot did not indicate skew in the
distribution of effect sizes, and/or the need to add more effect sizes on any side of the distribution to make it more symmetric. Together, these tests consistently suggest that publication bias was not a major concern in our sample.

**Results**

**Mediation Analysis**

Although a main effect between hierarchy and effectiveness is not a requirement for mediation to exist (Aguinis, Edwards, & Bradley, 2017), we nonetheless assessed this overall main effect in order to gain insight into whether the positive or the negative effect of hierarchy is more likely to win out at this point in the literature. Our results show a net negative point estimate in relation to team performance ($\rho = -.08$). The CI indicates that the true main effect size is unlikely to be zero (CI$_{95\%} = -.13, -.04$), indicating the presence of a negative effect of hierarchy on performance. Likewise, the findings also indicate a negative effect size estimate in relation to team viability ($\rho = -.11$), with the CI suggesting a true effect size is highly likely to be below zero despite the non-trivial uncertainty surrounding the point estimate (CI$_{95\%} = -.19, -.02$). Thus, we find that, at this point in the literature, hierarchy, on net, negatively impacts both team performance and viability.

We subsequently proceeded to analyze the mediated effects of hierarchy using meta-analytic structural equation modeling (MASEM) procedures outlined by Viswesvaran and Ones (1995). The first step in this analysis involves estimating true mean correlations for all pairs of variables that make up the conceptual model, and pooling these correlations into a matrix. To populate this matrix, we used the afore-mentioned Hedges and Olkin approach with random effects to calculate mean corrected correlations of hierarchy with each of the mediators and team effectiveness outcomes. Correlation estimates among the mediators and team effectiveness
outcomes were obtained from prior meta-analyses conducted by LePine and colleagues (2008) and Maynard, Mathieu, Gilson, O'Boyle, and Cigularov (2013). Pooling these meta-analytically derived correlations yielded the matrix displayed in Table 1. In the second step, we specified our mediation model – with hierarchy as the independent variable, coordination-enabling processes and conflict-enabling states as mediators, and team performance and viability as dependent variables – and used the correlation matrix as input to test this model in LISREL 9.1 (Jöreskog & Sörbom, 1996). Given that sample sizes differed across meta-analytical correlations, we used the harmonic mean as our sample size. Subjecting the correlation matrix to LISREL and testing our multiple mediation model yielded beta-coefficient estimates, standard errors, and CIs for each of the individual paths. We subsequently used the RMediation package (Tofighi & MacKinnon, 2011), which relies on the distribution-of-the-product method, to build 95% CIs around the mediated effect (cf. Courtright et al., 2015). Hypothesis 1 predicted that coordination-enabling processes would mediate the positive relationship between hierarchy and our two team effectiveness outcomes. Contrary to Hypothesis 1a, the results shown in Table 2 indicate that while coordination-enabling processes do mediate the hierarchy-performance relationship, and the mediated effect estimate is below-average in magnitude, the point estimate is negative rather than positive ($ab = -.04$), with a negative path running from hierarchy to coordination-enabling processes ($\beta = -.17$) and a positive path running from coordination-enabling processes to performance ($\beta = .21$). The CI indicates that while the there is a non-trivial degree of imprecision around the mediated effect estimate, the true mediated effect size parameter is likely to be negative ($\text{CI}_{95\%} = -.05$, -.02), and not positive as we hypothesized.

The same pattern of results was found for team viability. The findings indicate a below-average mediated effect of hierarchy on viability via coordination-enabling processes, but the
effect is negative rather than positive in nature ($ab = -.04$), with a negative path running from hierarchy to coordination-enabling processes (as already indicated above) and a positive path running from coordination-enabling processes to viability ($\beta = .23$). The CI indicates that the true mediated effect size is likely to be negative ($CI_{95\%} = -.06, -.02$), instead of the positive indirect effect we predicted. Thus, our results also fail to support Hypothesis 1b.

Hypothesis 2 proposed conflict-enabling states as a mediator of the negative relationship between hierarchy and team effectiveness. In support of Hypothesis 2a, we find a negative mediated effect estimate of hierarchy on performance via conflict-enabling states ($ab = -.02$), based on a positive path running from hierarchy to conflict-enabling states ($\beta = .17$) and a negative path running from conflict-enabling states to performance ($\beta = -.13$). The results also indicate that the true mediated effect parameter is below-average, but despite some imprecision in this estimate, it is likely to be negative ($CI_{95\%} = -.04, -.01$), indicating the presence of a negative indirect effect and support for Hypothesis 2a. Similarly, our results also show a negative indirect effect estimate of hierarchy on viability ($ab = -.03$), with a positive path from hierarchy to conflict-enabling states (as already indicated above) and a negative path from conflict-enabling states to viability ($\beta = -.18$). The effect is below-average, but the CI indicates that we can be confident that the true mediated effect size is likely to be negative ($CI_{95\%} = -.05, -.02$). These results provide support for Hypothesis 2b.

**Moderation Analysis**

To assess whether moderator analysis was warranted, we first assessed effect size heterogeneity across primary studies using two complementary indices: the $Q$-statistic and $I^2$ (Higgins & Thompson, 2002; Huedo-Medina, Sánchez-Meca, & Botella, 2006). Our analysis across the 52 effect size estimates for the hierarchy-performance relationship in our sample
showed the presence of considerable variation in effect sizes ($Q = 176.97, p = .00$) and suggests that most of this variation may be explained by between-study differences ($I^2 = 71.18$). We therefore proceeded to test our moderator hypotheses (H3-H5) by applying random effects, weighted least squares (WLS) meta-regression, which allows for examining the influence of continuous – as opposed to discrete – moderator variables, and examining the influence of multiple moderator variables simultaneously – as opposed to one at a time (Steel & Kammeyer-Mueller, 2002). While WLS meta-regression yields unstandardized beta coefficient estimates for each of the moderators ($B$), the effect size magnitude benchmark we rely on is expressed as a semi-partial correlation ($sr$). We therefore also converted our results into semi-partial correlations, using the formula provided by Cohen and Cohen (1983), in order to interpret the effect size magnitudes of our moderators in light of this benchmark.

**Team Task Contingencies.** We proposed in Hypothesis 3 that a) task interdependence; b) task ambiguity; or c) task complexity would moderate the positive relationship between hierarchy and team performance, such that the relationship would become stronger as the levels of these task characteristics increase. The results of meta-regression analysis failed to support Hypothesis 3a (see Table 3). While the coefficient estimate for the relationship of task interdependence with the hierarchy-performance correlation is above-average in magnitude ($sr = -.38$), it is negative rather than positive in nature ($B = -.07$). The CI around the estimate confirms that the true parameter coefficient is likely to be negative as opposed to positive ($CI_{95\%} = -.12, - .03$). These results thus contradict Hypothesis 3a. Hypothesis 3b regarding task ambiguity was supported. The results show a positive, above-average coefficient estimate for the relationship between task ambiguity and the correlation between hierarchy and team performance ($B = .14; sr = .45$), and suggest that we can be 95% confident that there is a positive, non-zero true
moderation effect in the population ($CI_{95\%} = .07, .21$). Hypothesis 3c regarding task complexity was not supported. The results yielded a positive effect ($B = .03$), but the estimate was below-average ($sr = .11$) and surrounded by relatively high levels of imprecision that include zero ($CI_{95\%} = -.03, .08$).

**Team Structure Contingencies.** Hypothesis 4 predicted that a) membership instability or b) skill differentiation would moderate the negative relationship between hierarchy and team performance, such that the relationship becomes stronger as the levels of these structural variables increase. As shown in Table 3, our results show support for Hypothesis 4a: membership instability was negatively related to the correlation between hierarchy and team performance ($B = -.05$), and the point estimate is above-average in magnitude ($sr = -.25$). While the upper limit of the CI approximates zero ($CI_{95\%} = -.10, -.004$), it does not include zero, indicating that while the estimate is surrounded by relatively high levels of imprecision, the true moderation effect is likely to be negative. The results also revealed support for Hypothesis 4b, showing a negative coefficient estimate for the relationship between skill differentiation and the correlation between hierarchy and team performance ($B = -.09$). The estimate is above-average in magnitude ($sr = -.36$) and the CI indicates that the true moderation effect parameter is likely to be negative ($CI_{95\%} = -.15, -.03$). These findings offer support for Hypothesis 4.

**Team Hierarchy Contingencies.** We proposed in Hypothesis 5 that a) hierarchy mutability or b) hierarchical form would moderate the negative relationship between hierarchy and performance, such that the relationship becomes stronger as mutability increases and hierarchy is based on inequality (as opposed to acyclicity). In support of Hypothesis 5a, we find a negative relationship between hierarchy mutability and the hierarchy-team performance correlation ($B = -.11$), with an above-average effect size magnitude ($sr = -.35$), and CIs that
suggest we can be confident that the true moderation effect in the population is negative ($CI_{95\%} = -0.19, -0.04$). Thus, Hypothesis 5a is supported. Contrary to Hypothesis 5b, the results indicate a negative coefficient estimate for acyclic hierarchies ($B = -0.48$) that was above-average in magnitude ($sr = -0.30$), suggesting that for such hierarchy forms the relationship between hierarchy and performance is considerably more negative than for inequitable hierarchies (whether steep or centralized). To gain further insight, we disaggregated inequality into two separate hierarchy forms (steepness and centralization) and employed a ‘shifting unit of analysis’ approach (Cooper, 2010), which involves computing separate correlations for each hierarchy form when hierarchy forms were reported within the same study (thereby increasing the number of acyclic studies to two). We then re-analyzed our regression model twice, once using steepness and once using centralization as the reference category. These analyses revealed that the population coefficient for acyclic hierarchies was likely to be considerably more negative than that for steep hierarchies ($B = -0.28; CI_{95\%} = -0.52, -0.04$), but unlikely to be different in magnitude from that of centralized hierarchies ($B = -0.18; CI_{95\%} = -0.42, 0.06$). While the CIs suggest a true difference in effect size between these hierarchy forms is likely to be present in the population ($CI_{95\%} = -0.83, -0.12$), the result was based on a restricted sample of only two studies containing an acyclic measure, making it hard to confidently interpret these results. Irrespective, Hypothesis 5b is not supported.

Together our seven moderators explain 43% of the variance in hierarchy-performance correlations across studies.

**Supplementary Analyses**

**Hierarchy Measurement.** Because the ways in which the authors in the hierarchy literature measure hierarchy vary quite dramatically, we examined whether hierarchy’s effects on
performance varied based on how hierarchy was measured. As seen in Table 4, the effect sizes do not markedly differ across the different measurements of intragroup hierarchy. Hierarchies measured on differences in expertise levels appear to be the most positive for team performance, while hierarchies based on formal positions in a team appear to be the most negative. However, given that these estimates are based on relatively small sample sizes, they may be susceptible to second-order sampling error (Schmidt & Hunter, 2015). We therefore caution readers to avoid over-interpreting potential differences across hierarchy measures reported in Table 4.

**Causality.** In our sample, some studies were causal (experimental) while others were correlational (field) in nature. To explore whether our findings across this broad sample are on par with only looking at the effects from purely causal studies, we checked whether the findings differed between the correlational and experimental studies using WLS meta-regression. We did not find meaningful differences between experimental studies (the reference category) and correlational studies in the effect of hierarchy on team effectiveness ($B = -.02$, CI$_{95\%} = -.07, .11$).

**Mediation Analyses.** To gain further insight into the generalizability of our mediation results, we complemented Viswesvaran and Ones’s (1995) fixed effects MASEM approach with a random effects (FIMASEM) approach recently developed by Yu, Downes, Carter, and O'Boyle (2016). This approach uses both meta-analytical correlations and standard deviations (reported in Table 1) to generate a large number of bootstrap samples on which the proposed mediation model is tested, thereby providing insight into the effect size heterogeneity associated with individual path coefficients and into the percentage of samples that yield an acceptable fit of the model as a whole. In short, the results of the FIMASEM analysis show path coefficients for individual and mediated paths that are directionally consistent with those from the fixed effects approach, as would be expected. More importantly, though, they also indicate that both the
individual path estimates and the negative dual-pathway model as a whole generalize to a majority of the bootstrap samples. These results thus provide further support for Hypotheses 2a-b and further evidence against Hypotheses 1a-b. More details can be found in the online supplements.

**Moderation Analyses.** While the advantage of WLS regression is that it accounts for shared variance among the moderators, estimating seven moderators simultaneously with only 52 studies could potentially produce unstable estimates due to statistical power issues. Indeed, this might explain why some of the findings did not support our hypotheses. To examine this possibility, we re-ran our moderator analysis twice, once for the subset of moderators that did support our hypotheses (i.e., task ambiguity, membership instability, skill differentiation, hierarchy mutability) and once for the subset of moderators that did not (i.e., task interdependence, task complexity, hierarchy form). This approach helps overcome issues of statistical power by increasing the moderator-to-study ratio, while still accounting for shared variance by including multiple moderators within each test. These analyses produced results very similar to our original analyses, in that they also supported the moderating roles of task ambiguity, skill differentiation, and hierarchy mutability, and also failed to support the moderating roles of task interdependence, task complexity, and hierarchy form. The only exception was the result for membership instability: the moderating effect supported in our original analysis was not supported when re-running the analysis on the moderator subset. Taken together, the consistent results strengthen confidence in the results from our original analysis and the stability of the estimates for nearly all of our moderators, including those that did not support our hypotheses. These analyses can be obtained from the second author.

**Discussion**
Our investigation of why and when hierarchy impacts team effectiveness represents a first conceptual and empirical integration of the different perspectives on hierarchy in teams, including the functionalist, conflict, and contingency perspectives. Together, the key variables and predictions associated with these perspectives and our empirical findings in this meta-analysis are indicative of an emergent, integrative conceptual framework of the way hierarchy operates in team settings (see Figure 1). We recognize that our meta-analytic approach is limited in the sense that our model is derived from an aggregation of ideas and results presented in prior research. We therefore highlight here how our review and findings are indicative of a theory of hierarchy in teams which can be more fully and robustly tested through future primary studies that examine the set of relationships we suggest may categorize the broader literature here.

Specifically, as the figure illustrates, in the field of research on hierarchy in teams, functionalist and conflict perspectives have painted very different pictures of the effects of hierarchy on team outcomes (positive versus negative). Additionally, different suggestions have arisen on the upstream contingencies which can reconcile these perspectives, and determine whether hierarchy sets into play performance-detracting conflict states or performance-enhancing coordinated action processes. By examining the upstream contingencies which determine the types of processes hierarchy sets into play in teams, researchers can develop the most targeted interventions to course-correct the effects of hierarchy at their root. Our meta-analytic examination provides a first direct comparison of the mechanisms and moderators proposed within these different perspectives, and our findings suggest that several key aspects of these perspectives portrayed in our figure are in need of further qualification and extension in future primary studies. For instance, our findings challenge the strength of the functionalist perspectives, add nuance to the conflict perspectives by identifying amplifying or ameliorating
contingencies, and parse out the most potent moderators from the less potent ones. These qualifications and extensions have important implications for the hierarchy literature moving forward, which we elaborate on in more detail below.

**Theoretical Implications and Avenues for Future Research**

The co-existence of the functionalist and conflict perspectives on hierarchy has given rise to the theoretical notion that hierarchy might act as a double-edged sword, capable of both helping and harming teams (Anderson & Willer, 2014). Our meta-analysis provide a first empirical test of this notion, and our results offer very little support for the functional perspective, but overwhelming support for the conflict perspective, showing an overall negative main effect of hierarchy across team effectiveness outcomes, a negative mediating effect of conflict-enabling states that generalizes across the majority of our bootstrapped samples, and support for three of the four moderators associated with the conflict perspective with above-average effect size magnitudes. This evidence suggests that, at this point in the literature, one edge of the sword’s blade – i.e., the conflict perspective – is considerably sharper than the other, and that there is a need to reconsider past assumptions on the functionality of hierarchy. Indeed, while the functionalist perspective has received much more theoretical attention, and is held as a common popular belief, our findings directly challenge this belief, suggesting that this belief may have been over-extended from just a few studies that found support for the positive mediating role of coordination (i.e., only three of the 54 studies in our meta-analysis found support for this: Boone & Hendricks, 2009; Halevy et al., 2012; Woolley et al., 2008), or from studies that did not directly test the impact of hierarchy on team task performance (i.e., Halevy et al., 2011; Tiedens & Fragale, 2003; Zitek & Tiedens, 2012). Our findings do not imply that hierarchy cannot be functional for teams, but they do suggest that the positive coordination
pathway proposed in the literature is of limited help in explaining why hierarchy enhances team effectiveness, and that there is need for future research to critically rethink the current functionalist perspective in the literature. For example, it may be that hierarchy generally is functional for team effectiveness, but may only spur the performance of certain individuals within the hierarchy (i.e., those on top).

By contrast, while the conflict perspective has, up until recently, received far less attention in the literature (for a recent review of this perspective, see Greer et al., 2017), our meta-analytical evidence suggests that this perspective offers a particularly potent explanation for how hierarchy operates in teams, and we extend thinking in this literature by incorporating the different mechanisms under the grouping of conflict-enabling states and by linking the theoretically connected moderators under the heading of conflict-susceptibilities. Our findings suggest that this perspective, and the extensions on it we make here, should receive more systematic attention in the future. In particular, we believe more work should be done in extending our understanding of the negative and nefarious ways in which hierarchy can drive performance-deterring conflicts in teams. For one example, research might focus on unpacking the micro-mediating mechanisms underlying the relationship between hierarchy and conflict, and identifying contingencies or interventions best suited to mitigate these mechanisms. To illustrate, consider Van Bunderen et al.’s (2017) suggestion that hierarchy relates to conflict because it leads members at different ranks to have different needs and interests. This suggests that settings which amplify different individual interests should bring out the most negative effects of hierarchy, and indeed we find that when skill differentiation is high (which should amplify members having different interests), hierarchy is most negative for performance. Interventions which can align members around collective interests, such as group-based rewards, would be
interesting to investigate in future research as one potential extension of the conflict perspective. As another example, we adopted a broad view of conflict-enabling states and coordination-enabling states in an effort to capture as many of the disparate mediators as we could and to provide an integrative look at the different positive or negative explanations of why hierarchy impacts team effectiveness. However, regarding the negative explains for instance, we do believe that hierarchy most likely relates to conflicts over ranks (i.e., status conflicts, Hays & Bendersky, 2015; power struggles, Greer & Van Kleef, 2010) or resources (i.e., process conflicts about the allocation of resources, De Wit, Greer, & Jehn, 2012). While data limitations did not allow us to test this idea, future research would therefore benefit from examining these specific types of conflict states in order to better understand whether the effects of hierarchy are bound to a certain form of conflict-enabling state, or generalize across all conflict-enabling states.

In addition to having implications and extensions for the broader functionalist and conflict perspectives on hierarchy, our results also extend emerging contingency perspectives (e.g., Anderson & Willer, 2014; Bunderson et al., 2016) on when hierarchy impacts teams. Specifically, in showing that some contingency factors can explain when hierarchy is particularly beneficial (e.g., high task ambiguity) while others explain when it is particularly harmful (e.g., high temporal instability), our findings suggest that contingency factors complement each other in explaining how hierarchy can be both helpful and harmful to teams. These findings point to the need to examine moderators in a more integrative manner, and shift attention to considering their joint impact, in order to more fully understand how hierarchy operates in teams. Conversely, our findings also imply that, in singling out individual moderators, prior research may have under- or overestimated the true impact of hierarchy within their particular research context (e.g., Bunderson et al., 2016; Ronay et al., 2012), and that they might have arrived at
different or more nuanced conclusions if they had simultaneously accounted for multiple contingencies that might have been at work. As such, our integrative analysis extends the contingency approach advocated by others (e.g., Anderson & Willer, 2014; Bunderson et al., 2016) by pointing out the need for considering integrating multiple complementary contingencies at a time. Specifically, in light of the limited evidence for the functional contingencies and convincing support for the conflictual contingencies, we identify understanding when the former outweigh the latter as an important direction for future work.

Beyond the general recommendation of examining multiple contingency factors, our findings also are suggestive of which moderators should be included in future research by providing insight into which moderators are the most powerful (i.e., task ambiguity, skill differentiation, membership instability, and hierarchy mutability) and which are the least (e.g., task complexity). The fact that the most powerful moderators offer unique predictive validity above and beyond the other moderators in our model attests to the robustness of their impact. For example, our results attest to the robustness of Hays and Bendersky’s (2015) recent finding on hierarchy mutability, suggesting that when hierarchies are more mutable, they are more likely to be contested and harm performance. Similarly, our results on team structures susceptible to conflict (high membership instability and/or skill differentiation) support theorizing by Gruenfeld and Tiedens (2010) that suggests that hierarchy and homogeneity are most effective when combined, and that teams can descend into silos and conflicts when only vertically or only horizontally differentiated. These first three moderators, and the robust support for them, can be explained in part due to their centrality within the conflict perspective – when teams have structures such as these, which leave teams easy pretty to conflict, hierarchies are likely to backfire for team performance. We did find one moderator which offered a ray of hope for
functional accounts of hierarchy – task ambiguity. By showing that when tasks are highly unclear, hierarchy does prove to be beneficial, task ambiguity may provide a useful lens to better understand how and why hierarchy could still be useful for teams, such as by bringing order to chaos and reducing uncertainty.

For the less supported moderators in our meta-analysis, it suggests that the effect size magnitude and direction of some moderators examined in past research may have been overstated or misrepresented. Notably, contrary to Bunderson and colleagues (2016), we actually found that acyclical hierarchies were more negative for team performance than inequitable hierarchies. We note that this finding should be interpreted with caution, as this comparison was based on only a small number of studies using acyclical hierarchy measures. Nonetheless, our results clearly signal the need to re-consider and explore how different forms of hierarchy, such as the forms introduced by Bunderson et al. (2016) as well as other facets of hierarchy, may differentially shape the hierarchy-performance relationship. Our predictions regarding task complexity were not supported, and for task interdependence, we even found that it enhanced the team performance deterrents of hierarchy. In retrospect, the latter finding might be understood in light of conflict theories– in loosely interdependent teams, members may not interact as much and thereby have less chance for conflict. In such situations, hierarchies may be less contested and more likely to help performance though as this literature progresses further work identifying the contingencies that elicit positive effects of hierarchy will be helpful.

Together, our findings challenge past contingency perspectives which have focused on single moderators, and extend past work by showing which combination of moderators may work best in combination – task ambiguity, which can explain the bright side of hierarchy, and conflict-susceptibility moderators, such as membership instability, skill differentiation, and
hierarchy mutability. These findings challenge the strength of the functionalist perspective, and add nuance to the conflict perspective by identifying the contingencies most relevant to amplifying or ameliorating the more negative effects of hierarchy in teams and thus help us interpret how to move forward in theorizing about the nature and effects of hierarchical differentiation in teams.

Methodological Implications

Our meta-analysis and emergent theory also offer considerations for the methodological designs of future primary studies. First, as we depict in our theoretical model, past research has proposed that hierarchy can benefit performance via coordination or hurt performance via conflict. When we incorporate both of these pathways in one model, we find that the negative pathway wipes out any benefits from the positive pathway. This suggests that future empirical studies would benefit from simultaneously examining both the positive as well as negative pathways by which hierarchy affects performance to understand the net overall effects hierarchy will have on performance. This could include examining conflict and coordination, or conflict and other potentially more powerful positive forces by which hierarchy could impact performance, such as shared mental models or role clarity.

Second, and relatedly, we discuss the benefits here of examining multiple moderating factors in conjunction, as this allows the direct comparison of different contingency factors. Because our conclusions are limited to inference from others’ primary data here, we encourage future research to continue to compare the predictive ability of variables from the two different moderator categories that we induce in our model here (task coordination demands and intra-team conflict susceptibilities). For example, research could compare whether task complexity or intra-team diversity better determine whether hierarchy helps or hurts team performance.
Third, when designing future primary studies, we believe multi-method research is important. We did not find large differences here based on how hierarchy was measured, and especially not in lab or field differences, and continuing to examine hierarchy in both lab and field settings can help rule out potential concerns about both generalizability in lab studies as well as third variable problems in field studies.

Fourth, when designing future primary studies, we note that while there do not appear to be large effects based on the type of hierarchy being measured, some hierarchies (i.e., expertise) may have more positive effects than others (i.e., positional). Therefore, being intentional about the base of hierarchy one measures and making sure the theory is explicitly tied to the base being measured is important and may yield more nuanced and precise knowledge in this area as we go forward. We noted that historically many papers would not align the base of hierarchy being measured with the theory, theorizing about status hierarchies and measuring influence hierarchies, or theorizing about power hierarchies and then measuring positional hierarchies which may reflect both power and status differences in teams. More precision in construct clarity thus can help the field move forward.

**Limitations and Future Opportunities**

In addition to setting out future areas in need of conceptual attention in future research, our meta-analysis also highlights issues to keep in mind in future empirical hierarchy research. First, while the effect size estimates for the supported moderators are all above-average, the effect sizes for our main and mediated effects were below-average. There are several reasons why this might be the case. First, compared to proximal outcomes (e.g., emergent states, behavioral processes), team performance represents a very distal outcome to team hierarchy that is more likely to be impacted by exogenous factors that are not under the direct control of the
team (e.g., environmental rivalry, turbulence, etc.; Beal, Cohen, Burke, & McLendon, 2003), and therefore typically yield smaller effect size magnitudes (De Wit et al., 2012). Indeed, our findings regarding impact of hierarchy on the two mediators versus performance are consistent with this pattern. Second, the vast majority of the team performance measures were measured objectively (75% of studies) rather than subjectively, which tends to result in smaller effect size magnitudes due to a combination of reduced bias and decreased content adequacy (Mesmer-Magnus & DeChurch, 2009). A third reason is that there was considerable effect size heterogeneity across the primary studies in our sample, which dampened the overall mean effect size estimates of the main and mediated effect size estimates. While we accounted for this heterogeneity in the main effects via our moderator analyses, data limitations (i.e., small number of effect sizes between hierarchy and the mediators) prevented us from accounting for heterogeneity in the mediated effects (i.e., integratively testing mediation and moderation). Indeed, no team-level meta-analysis, to our knowledge, has ever had the sample size to test a full moderated mediation model, and other team-level meta-analysis, similar to ours, have therefore tested moderation and mediation separately (e.g., Maynard, Mathieu, Gilson, O’Boyle, & Cigularov, 2013; Nicolaides, LaPort, Chen, Tomassetti, Weis, Zaccarro, & Cortina, 2013; Stajkovi, Lee, & Nyberg, 2009; Whitman, Va Rooy, & Viswesvaran, 2010). Future work will thus benefit from larger sample sizes to detect modest effect sizes and to perform a more integrative analysis of mediated moderation which can account for mediated effect size heterogeneity (Aguinis et al., 2017).

Second, we note also limitations regarding sample size and, hence, statistical power. While our cumulative sample sizes do offer considerable statistical power for main and mediated effects, there is considerably less power for our moderating effects (although our supplementary
analyses do show most results hold when analyzing with more power). As such, we were only able to examine moderators of the impact of hierarchy on team performance, and not on our proposed mediators or team viability. While we expect the moderators in the literature to likely operate upstream in our model (as shown in Figure 1, where we highlight that the literature suggests the moderators to moderate the upstream link between hierarchy and team processes, rather than the downstream link between processes and outcomes; see for example Bunderson et al., 2016), future research which can have the sample size to fully test moderated mediation models will be important. Additionally, the number of effect size estimates for certain moderators was rather small (e.g., the number of acyclic studies), as was the number of effect size estimates for the relationships between hierarchy and the two mediators, making those results susceptible to second-order sampling error (Schmidt & Hunter, 2015). Future work may thus benefit from exploring these less represented moderators (e.g., acyclicity), mediators (e.g., conflict, coordination), and dependent variables in the hierarchy literature (e.g., viability).

Third, we noted when conducting this meta-analysis that a considerable portion of the primary studies \(k = 30\) relied on field data. While we do not find the conclusions of those studies to differ from those conducted in experimental settings in our supplementary analysis, we would strongly encourage future research on hierarchy to focus more strongly on causal methods. Relatedly, while the literature to this point has largely focused on the effects of hierarchy on team outcomes (e.g., Bunderson et al., 2016), team outcomes can in turn affect team inputs. We therefore strongly encourage work that not only examines the causal impact of hierarchy on team outcomes, but also examines the causal impact of team outcomes on the formation and stability of team hierarchies and recursive models of hierarchy.

Finally, we intentionally adopted a very broad definition of hierarchy here – of vertical
differences in socially valued resources, yet as we note in this definition and explore with our construct-related moderators, there is much left to be understood about the different ways hierarchies are constructed in teams in terms of the types of resources on which hierarchies are based as well as the form or structure of vertical differences. In terms of the types of resources, we focused here on resources most likely to be socially valued in small work teams – directly task-related resources (Berger et al., 1972; Bunderson, 2003), but other resources, including race or gender, may also potentially affect work team interactions, and would benefit from future research. Additionally, within work-related differences, we did not find many significant differences between most types of resources included in our analyses (results available from second author upon request), but others have found resources such as power and status to differ (e.g., Hays & Bendersky, 2015). Thus, future research would benefit from deeper understanding of the resources underlying hierarchy. In terms of the form of hierarchy, we did not find form to matter here, but we only looked at one conceptualization of form (Bunderson et al., 2016), and one key form in that typology (acylicity) was not well-represented in the literature yet. Future research could benefit from better understanding acyclicity, as well as the other potential different forms of vertical differences, such as the impact of the number or breadth of layers in a hierarchy. Such studies could also perhaps provide more insight into when and how the functional perspective on hierarchy may still apply.

**Conclusion**

In sum, our meta-analytic investigation of why and when hierarchy impacts team effectiveness helps parse the widespread literature that has arisen on the topic of hierarchy and identifies priorities for future research. Our study challenges several prior key propositions in the field, including the functionalist perspective on hierarchy’s effectiveness. Our study in contrast
extends the conflict perspective on hierarchy, as we show that the relationship between hierarchy and team effectiveness is negative on net, that this effect is mediated by conflict-enabling states, and exacerbated in teams seemingly susceptible to conflict (i.e. teams with temporal instability, skill differentiation, and/or mutable hierarchies). Our study therefore qualifies the functional perspective on hierarchy and shows the need to further unpack and elaborate on the conflict perspective on hierarchy.
References*

*Papers in the meta-analytic dataset are marked with a single asterisk. Prior team meta-analyses included in our effect size benchmark (seen in Appendix B) are marked with double asterisks.


*Cao, J. Top management team power inequality: The moderating roles of environmental dynamics. Working paper.


DeDreu, C. K. W., & Gelfand, M. J. (2008). Conflict in the workplace: Sources, functions, and
dynamics across multiple levels of analysis. In C. K. W. De Dreu & M. J. Gelfand (Eds.),
The psychology of conflict and conflict management in organizations (3–54). New York, NY:
Erlbaum.

De Dreu, C. K., & Weingart, L. R. (2003). Task versus relationship conflict, team performance,
and team member satisfaction: a meta-analysis. Journal of Applied Psychology, 88(4), 741-
749. doi.org/10.1037/0021-9010.88.4.741

De Jong, B. A., Dirks, K. T., & Gillespie, N. (2016). Trust and team performance: A meta-
analysis of main effects, moderators, and covariates. Journal of Applied Psychology, 101(8),
1134-1150. doi: 10.1037/apl0000110

The mediating role of reflexivity, monitoring, and effort. Academy of Management Journal,

decisions in social dilemmas: Justification pressures and tacit coordination under
doi: 10.1177/0146167207307490

De Wit, F. R. C., Greer, L.L., & Jehn, K.A. (2012). The paradox of intragroup conflict: A meta-

Devine, D. J., & Philips, J. L. (2001). Do smarter teams do better a meta-analysis of cognitive
ability and team performance. Small Group Research, 32(5), 507-532. doi:
10.1177/104649640103200501


Galinsky, A.D., Chou, E.Y., Halevy, N., & Van Kleef, G.A. (2012). The far-reaching effects of power: At the individual, dyadic, and group levels. In M.A. Neale & E.A. Mannix (Eds.)


doi:10.1037/a0020373


206. doi.org/10.1037/1082-989X.11.2.193


**Maynard, M. T., Mathieu, J. E., Gilson, L. L., O’Boyle, E. H., & Cigularov, K. P. (2013). Drivers and outcomes of team psychological empowerment a meta-analytic review and
model test. *Organizational Psychology Review, 3*(2), 101-137. doi:
10.1177/2041386612456868

doi:10.1111/j.1467-6494.1953.tb01810.x

of how group identification affects status in work groups. *European Journal of Work and
Organizational Psychology, 1*-12. doi: 10.1080/1359432X.2015.1028378

organizational identity affects team functioning: The identity instrumentality


inclusiveness and diversity in teams: How leader inclusiveness affects performance through
status and team identity. *Human Resource Management, 54*(2), 217-239. doi:
10.1002/hrm.21658

*Nederveen Pieterse, A., Hollenbeck, J. R., van Knippenberg, D., Spitzmuller, M., Karam, E. P.,
Erasmus University Rotterdam, working paper.

Ng, T. H. W. (in press). Transformational leadership and performance outcomes: Multiple


*Van Dijk, H., Meyer, B., & Van Engen, M. L. (2015). Reconsidering the diversity-


### Table 1. Meta-Analytic Correlations and Standard Deviations Matrix

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Hierarchy</td>
<td></td>
<td>0.17&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.26&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.15&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.12&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>2 Coordination</td>
<td>-0.17 (13, 960)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.40&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.14&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.17&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>3 Conflict</td>
<td>0.17 (13, 1052)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.6 (8, 529)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.20&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.03&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>4 Team performance</td>
<td>-0.08 (52, 13781)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.29 (30, 2012)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-0.26 (32, 2049)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.19&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>5 Team viability</td>
<td>-0.11 (8, 540)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.34 (11, 902)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-0.32 (13, 1039)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.32 (21, 1512)&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
</tr>
</tbody>
</table>

*Note:* results below the diagonal are reported as \(r(k, N)\), where \(r\) = corrected correlation coefficient; \(k\) = number of independent studies; and \(N\) = cumulative sample size. Results above the diagonal are the standard deviations for the corrected correlations. \(a\) = correlations and standard deviations based on our own sample and analysis; \(b\) = correlations and standard deviations taken from LePine et al. (2008); \(c\) = correlations and standard deviations taken from Maynard et al. (2013). All meta-analytic results were generated using a random effects model.
### Table 2. Results of (Fixed Effects) MASEM analysis

<table>
<thead>
<tr>
<th>Individual paths</th>
<th>β</th>
<th>SE</th>
<th>CI 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hierarchy → coordination</td>
<td>-0.17</td>
<td>0.03</td>
<td>-0.23 -0.11</td>
</tr>
<tr>
<td>Coordination → performance</td>
<td>0.21</td>
<td>0.03</td>
<td>0.15 0.27</td>
</tr>
<tr>
<td>Coordination → viability</td>
<td>0.23</td>
<td>0.03</td>
<td>0.17 0.29</td>
</tr>
<tr>
<td>Hierarchy → conflict</td>
<td>0.17</td>
<td>0.03</td>
<td>0.11 0.23</td>
</tr>
<tr>
<td>Conflict → performance</td>
<td>-0.13</td>
<td>0.03</td>
<td>-0.19 -0.08</td>
</tr>
<tr>
<td>Conflict → viability</td>
<td>-0.18</td>
<td>0.03</td>
<td>-0.24 -0.12</td>
</tr>
<tr>
<td>Harmonic sample size</td>
<td>1049</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mediated paths</th>
<th>ab</th>
<th>SE</th>
<th>CI 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hierarchy → coordination → perf</td>
<td>-0.04</td>
<td>0.01</td>
<td>-0.05 -0.02</td>
</tr>
<tr>
<td>Hierarchy → coordination → viab</td>
<td>-0.04</td>
<td>0.01</td>
<td>-0.06 -0.02</td>
</tr>
<tr>
<td>Hierarchy → conflict → perf</td>
<td>-0.02</td>
<td>0.01</td>
<td>-0.04 -0.01</td>
</tr>
<tr>
<td>Hierarchy → conflict → viab</td>
<td>-0.03</td>
<td>0.01</td>
<td>-0.05 -0.02</td>
</tr>
</tbody>
</table>

Note: β = path coefficient; SE = standard error; CI = confidence interval; ab = indirect effect estimate
Table 3. Results of the Meta-Regression Analysis

<table>
<thead>
<tr>
<th>Moderators</th>
<th>$B$</th>
<th>$SE$</th>
<th>CI 95%</th>
<th>sr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task interdependence</td>
<td>-0.07</td>
<td>0.02</td>
<td>-0.12</td>
<td>-0.38</td>
</tr>
<tr>
<td>Task ambiguity</td>
<td>0.14</td>
<td>0.04</td>
<td>0.07</td>
<td>0.21</td>
</tr>
<tr>
<td>Task complexity</td>
<td>0.03</td>
<td>0.03</td>
<td>-0.03</td>
<td>0.08</td>
</tr>
<tr>
<td>Membership instability</td>
<td>-0.05</td>
<td>0.02</td>
<td>-0.10</td>
<td>-0.25</td>
</tr>
<tr>
<td>Skill differentiation</td>
<td>-0.09</td>
<td>0.03</td>
<td>-0.15</td>
<td>-0.36</td>
</tr>
<tr>
<td>Hierarchy mutability</td>
<td>-0.11</td>
<td>0.04</td>
<td>-0.19</td>
<td>-0.35</td>
</tr>
<tr>
<td>Hierarchy form: acyclicity</td>
<td>-0.48</td>
<td>0.18</td>
<td>-0.83</td>
<td>-0.30</td>
</tr>
<tr>
<td>Hierarchy form: mixed</td>
<td>0.01</td>
<td>0.10</td>
<td>-0.18</td>
<td>0.20</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.48</td>
<td>0.20</td>
<td>0.09</td>
<td>0.28</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.43</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: $k = 52$. $B$ = unstandardized regression coefficient; $SE$ = standard error; $CI$ = confidence interval; $sr$ = semi-partial correlation coefficient; $R^2$ = variance explained by the model; reference category for hierarchy form: inequality.
<table>
<thead>
<tr>
<th>Hierarchy Measure</th>
<th>k</th>
<th>r</th>
<th>CI</th>
<th>95%</th>
<th>Typical Measurement</th>
<th>Example Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expertise Hierarchy</td>
<td>3</td>
<td>0.07</td>
<td>-0.11</td>
<td>0.24</td>
<td>Measuring the dispersion of expertise depth within a team, such as via the coefficient of variation or standard deviation of member expertise levels in the team</td>
<td>Huang &amp; Cummings (2011), Trzebiatowski &amp; Trevor (2016), Woolley et al. (2008)</td>
</tr>
<tr>
<td>Influence Hierarchy</td>
<td>15</td>
<td>-0.09</td>
<td>-0.17</td>
<td>0.02</td>
<td>Measuring the dispersion of influence in the team, such as by round robin measures of influence and then taking the coefficient of variation of member influence, or also as measured by variation in speaking time</td>
<td>Bunderson et al. (2016), Cummings &amp; Cross (2003), Sauer &amp; Kauffield (2013)</td>
</tr>
<tr>
<td>Negotiation BATNAS</td>
<td>7</td>
<td>-0.10</td>
<td>-0.22</td>
<td>0.01</td>
<td>Manipulating the alternatives to agreement individuals have in team negotiations, such that one person has a better alternative than the other</td>
<td>Giebels et al. (2000), Mannix (1993), Wolfe et McGinn (2005)</td>
</tr>
<tr>
<td>Perceived Hierarchy</td>
<td>4</td>
<td>-0.19</td>
<td>-0.62</td>
<td>0.33</td>
<td>Manipulating the sense of power of a leader or measuring the perceptual hierarchical climate in a team</td>
<td>Curseu &amp; Sau (2013), Perry (2016); Tost et al. (2013)</td>
</tr>
<tr>
<td>Positional Hierarchy</td>
<td>17</td>
<td>-0.09</td>
<td>-0.16</td>
<td>-0.01</td>
<td>Measuring or manipulating the presence of different formal ranks within a team, such as the number of corporate levels in the team or manipulating the presence of a formal leader or not in a team</td>
<td>Greer &amp; Van Kleef (2010); Hambrick et al. (2015), Van Bunderen et al. (2017)</td>
</tr>
<tr>
<td>Status Hierarchy</td>
<td>6</td>
<td>-0.09</td>
<td>-0.22</td>
<td>0.04</td>
<td>Measuring the dispersion of status in the team, such as by round robin measures of status and then taking the coefficient of variation of member status, or by taking the gini coefficient of member status markers such as board memberships</td>
<td>Greer et al. (2016), He &amp; Huang (2011), Thylefors (2004)</td>
</tr>
</tbody>
</table>

Note: $k = \text{number of independent studies}; \ r = \text{corrected correlation coefficient}; \ CI = \text{confidence interval}.$
Figure 1. Conceptual model of suggested integration of different theoretical perspectives on why and when hierarchy impacts team effectiveness.
### Appendix A. Overview of sample size, reliabilities, and moderators for all studies included in meta-analysis

<table>
<thead>
<tr>
<th>Article</th>
<th>Sample Size</th>
<th>Reliabilities</th>
<th>Moderators</th>
<th>Team Inter.</th>
<th>Task Ambi.</th>
<th>Task Compl.</th>
<th>Member Instability</th>
<th>Skill Diff.</th>
<th>Hierarchy Mutability</th>
<th>Hierarchy Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Becker &amp; Baloff (1969)</td>
<td>36</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>-</td>
<td>M</td>
<td>H</td>
<td>M</td>
<td>ML H S</td>
</tr>
<tr>
<td>Bona (2007)</td>
<td>54</td>
<td>1.00</td>
<td>1.00</td>
<td>-0.21</td>
<td>0.86</td>
<td>0.83</td>
<td>0.83</td>
<td>0.09</td>
<td>MH</td>
<td>M M L M C</td>
</tr>
<tr>
<td>Boone &amp; Hendriks (2009)</td>
<td>33</td>
<td>0.95</td>
<td>0.40</td>
<td>0.07</td>
<td>1.00</td>
<td>0.32</td>
<td>-</td>
<td>-</td>
<td>M</td>
<td>ML H L H ML S</td>
</tr>
<tr>
<td>Brett et al. (1996)</td>
<td>163</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>0.04</td>
<td>-</td>
<td>-</td>
<td>L</td>
<td>L ML H ML S</td>
</tr>
<tr>
<td>Bunderson (2003)</td>
<td>35</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>0.95</td>
<td>0.02</td>
<td>-</td>
<td>H</td>
<td>L</td>
<td>L ML M ML C</td>
</tr>
<tr>
<td>Bunderson et al. (2016)</td>
<td>75</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>0.52</td>
<td>0.09</td>
<td>0.84</td>
<td>0.50</td>
<td>0.09</td>
<td>M M M A Mix</td>
</tr>
<tr>
<td>Cantimur et al. (2015a)</td>
<td>57</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>0.72</td>
<td>-0.37</td>
<td>-</td>
<td>-</td>
<td>M</td>
<td>M M L M A</td>
</tr>
<tr>
<td>Cantimur et al. (2015b)</td>
<td>72</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>0.63</td>
<td>-0.03</td>
<td>0.85</td>
<td>-0.02</td>
<td>-</td>
<td>MH M L M S</td>
</tr>
<tr>
<td>Cao (2015)</td>
<td>5,167</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>-0.04</td>
<td>-</td>
<td>-</td>
<td>M</td>
<td>ML H ML H ML S</td>
</tr>
<tr>
<td>Curescu &amp; Sari (2015)</td>
<td>87</td>
<td>-</td>
<td>1.00</td>
<td>-0.13</td>
<td>-</td>
<td>-</td>
<td>0.83</td>
<td>-0.02</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Doornenbal et al. (2015)</td>
<td>127</td>
<td>1.00</td>
<td>0.40</td>
<td>-0.24</td>
<td>1.00</td>
<td>0.09</td>
<td>-</td>
<td>MH</td>
<td>H</td>
<td>ML MH L H S</td>
</tr>
<tr>
<td>Foo et al., (2006)</td>
<td>51</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>-0.17</td>
<td>0.73</td>
<td>0.06</td>
<td>H</td>
<td>H H M M H S</td>
</tr>
<tr>
<td>Gajendran (2009)</td>
<td>168</td>
<td>1.00</td>
<td>0.65</td>
<td>-0.19</td>
<td>0.89</td>
<td>0.02</td>
<td>-</td>
<td>-</td>
<td>MH</td>
<td>MH MH M H S</td>
</tr>
<tr>
<td>Giebels et al. (2000)</td>
<td>71</td>
<td>1.00</td>
<td>-</td>
<td>0.62</td>
<td>0.13</td>
<td>1.00</td>
<td>-0.23</td>
<td>-</td>
<td>L</td>
<td>L ML H ML ML S</td>
</tr>
<tr>
<td>Greer, Dannals, &amp; Rao (2016)</td>
<td>80</td>
<td>1.00</td>
<td>0.78</td>
<td>-0.19</td>
<td>1.00</td>
<td>-0.09</td>
<td>-</td>
<td>-</td>
<td>H</td>
<td>MH H M M H C</td>
</tr>
<tr>
<td>Greer &amp; Van Kleef (2010) Study 1</td>
<td>42</td>
<td>1.00</td>
<td>-</td>
<td>0.20</td>
<td>-0.01</td>
<td>0.45</td>
<td>-0.13</td>
<td>-</td>
<td>M</td>
<td>ML M ML MH ML S</td>
</tr>
<tr>
<td>Greer &amp; Van Kleef (2010) Study 2</td>
<td>40</td>
<td>1.00</td>
<td>-</td>
<td>0.21</td>
<td>0.25</td>
<td>1.00</td>
<td>-0.09</td>
<td>-</td>
<td>ML</td>
<td>L ML H ML ML S</td>
</tr>
<tr>
<td>Halebian &amp; Finkelstein (1993)</td>
<td>235</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>-0.05</td>
<td>-</td>
<td>-</td>
<td>M</td>
<td>ML H L H ML S</td>
</tr>
<tr>
<td>Article</td>
<td>n</td>
<td>$r_{yx}$</td>
<td>$r_{yy}$</td>
<td>$r_{xy}$</td>
<td>$r_{yy}$</td>
<td>$r_{xy}$</td>
<td>$r_{yy}$</td>
<td>$r_{xy}$</td>
<td>Team Inter.</td>
<td>Task Ambi.</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Hambrick et al. (2015)</td>
<td>365</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>-0.03</td>
<td>-</td>
<td>M</td>
<td>ML</td>
</tr>
<tr>
<td>Haslam et al. (1998)</td>
<td>46</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>-0.21</td>
<td>-</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>He &amp; Huang (2011)</td>
<td>2,157</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>-0.04</td>
<td>-</td>
<td>M</td>
<td>ML</td>
</tr>
<tr>
<td>Hulpia et al. (2012)</td>
<td>46</td>
<td>-0.31</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.91</td>
<td>-0.22</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Lemoine et al. (2015)</td>
<td>141</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>-0.07</td>
<td>-</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>Mannix (1993)</td>
<td>36</td>
<td>1.00</td>
<td>-</td>
<td>0.38</td>
<td>1.00</td>
<td>-0.54</td>
<td>-</td>
<td>-</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>Mannix &amp; Neale (1993)</td>
<td>675</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>-0.16</td>
<td>-</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>Martins et al. (2013)</td>
<td>196</td>
<td>0.62</td>
<td>-0.02</td>
<td>1.00</td>
<td>0.21</td>
<td>-</td>
<td>-</td>
<td>MH</td>
<td>H</td>
<td>MH</td>
</tr>
<tr>
<td>McCurdy &amp; Eber (1952)</td>
<td>21</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>0.02</td>
<td>-</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>Meussen &amp; Van Dijk (2015)</td>
<td>33</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>-0.11</td>
<td>-</td>
<td>MH</td>
<td>H</td>
</tr>
<tr>
<td>Mitchell et al. (2015)</td>
<td>75</td>
<td>0.89</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.94</td>
<td>-0.46</td>
<td>-</td>
<td>MH</td>
<td>L</td>
</tr>
<tr>
<td>Nederveen et al. (2014)</td>
<td>57</td>
<td>0.63</td>
<td>-0.09</td>
<td>-</td>
<td>1.00</td>
<td>-0.10</td>
<td>-</td>
<td>-</td>
<td>H</td>
<td>L</td>
</tr>
<tr>
<td>Patel &amp; Cooper (2014)</td>
<td>1,904</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>-0.19</td>
<td>-</td>
<td>M</td>
<td>ML</td>
</tr>
<tr>
<td>Perry (2014)</td>
<td>80</td>
<td>0.72</td>
<td>0.57</td>
<td>0.19</td>
<td>0.91</td>
<td>-0.23</td>
<td>1.00</td>
<td>-0.24</td>
<td>M</td>
<td>L</td>
</tr>
<tr>
<td>Pinkley et al. (1994)</td>
<td>142</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>0.06</td>
<td>-</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>Ren et al. (2015)</td>
<td>148</td>
<td>1.00</td>
<td>-</td>
<td>0.06</td>
<td>0.87</td>
<td>-0.04</td>
<td>-</td>
<td>-</td>
<td>MH</td>
<td>ML</td>
</tr>
<tr>
<td>Ronay et al. (2012)</td>
<td>46</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>0.51</td>
<td>-</td>
<td>M</td>
<td>H</td>
</tr>
<tr>
<td>Ruik &amp; Galaskiewicz (2000)</td>
<td>39</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>-0.13</td>
<td>-</td>
<td>M</td>
<td>L</td>
</tr>
<tr>
<td>Sauer &amp; Kauffeld (2013)</td>
<td>54</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>0.81</td>
<td>-0.27</td>
<td>-</td>
<td>-</td>
<td>H</td>
<td>ML</td>
</tr>
<tr>
<td>Small &amp; Rentsch (2010)</td>
<td>60</td>
<td>0.90</td>
<td>0.86</td>
<td>0.62</td>
<td>1.00</td>
<td>-0.17</td>
<td>-</td>
<td>-</td>
<td>MH</td>
<td>M</td>
</tr>
<tr>
<td>Article</td>
<td>n</td>
<td>$r_{xx}$</td>
<td>$r_{xy}$</td>
<td>$r_{yy}$</td>
<td>$r_{xy}$</td>
<td>$r_{yy}$</td>
<td>$r_{xy}$</td>
<td>$r_{yy}$</td>
<td>$r_{xy}$</td>
<td>$r_{yy}$</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>----</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>Sondak &amp; Bazerman (1991)</td>
<td>144</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td>-0.09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tarakei et al. (2016) Study 1</td>
<td>75</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td>0.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tarakei et al. (2016) Study 2</td>
<td>39</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.39</td>
<td>-0.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thylefors (2012)</td>
<td>62</td>
<td>0.91</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>0.99</td>
<td>-0.08</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Tost et al. (2013) Study 1</td>
<td>20</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>1.00</td>
<td>-0.58</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Tost et al. (2013) Study 2</td>
<td>48</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>1.00</td>
<td>-0.29</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Tost et al. (2013) Study 3</td>
<td>38</td>
<td>1.00</td>
<td>0.44</td>
<td>-0.31</td>
<td>-</td>
<td></td>
<td>1.00</td>
<td>-0.39</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Van Bunderen et al. (2015)</td>
<td>85</td>
<td>1.00</td>
<td>-</td>
<td>0.96</td>
<td>0.23</td>
<td></td>
<td>1.00</td>
<td>-0.09</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Van der Heijden et al. (2009)</td>
<td>20</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>1.00</td>
<td>0.69</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Van der Vegt et al. (2010)</td>
<td>46</td>
<td>1.00</td>
<td>0.62</td>
<td>0.00</td>
<td>-</td>
<td></td>
<td>0.90</td>
<td>0.14</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Van Dijk et al. (2015)</td>
<td>96</td>
<td>1.00</td>
<td>1.00</td>
<td>-0.09</td>
<td>-</td>
<td></td>
<td>1.00</td>
<td>0.05</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Van Dijk et al. (2015)</td>
<td>47</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>1.00</td>
<td>-0.07</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Wellman (2013) Study 1</td>
<td>87</td>
<td>1.00</td>
<td>0.38</td>
<td>-0.02</td>
<td>0.91</td>
<td>-0.04</td>
<td>0.81</td>
<td>-0.27</td>
<td>0.34</td>
<td>-0.05</td>
</tr>
<tr>
<td>Wellman (2013) Study 2</td>
<td>60</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>0.85</td>
<td>-0.14</td>
<td>0.89</td>
<td>-0.25</td>
<td>0.87</td>
<td>-0.05</td>
</tr>
<tr>
<td>Wolfe &amp; McGinn (2005)</td>
<td>62</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td>1.00</td>
<td>0.03</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Woolley et al. (2008)</td>
<td>41</td>
<td>1.00</td>
<td>1.00</td>
<td>0.15</td>
<td>-</td>
<td></td>
<td>1.00</td>
<td>-0.13</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** $n =$ number of groups; $r_{xx} =$ reliability hierarchy measure; $r_{xy} =$ observed correlation between hierarchy and dependent variable; $r_{yy} =$ reliability dependent variable; L = Low; ML = Moderately Low; M = Moderate; MH = Moderately High; H = High; C = Centralization; S = Steepness; A = Acyclicity; Mix = Mixed