Special Issue:
Bayesian Probability and Statistics
in Management Research

Journal of Management
Vol. 41 No. 2, February 2015 574–606
DOI: 10.1177/0149206314539351

Social Influence Interpretation of Interpersonal Processes and Team Performance Over Time Using Bayesian Model Selection

Alan R. Johnson
EMLYON Business School
Rens van de Schoot
Utrecht University
North-West University
Frédéric Delmar
Lund University
William D. Crano
Claremont Graduate University

The team behavior literature is ambiguous about the relations between members' interpersonal processes—task debate and task conflict—and team performance. From a social influence perspective, we show why members' interpersonal processes determine team performance over time in small groups. Together, over time, dissenting in-group minorities who share information (via debate) with majorities, who selectively engage with them to consider their alternative proposals (via conflict), can improve their team performance (via innovation). The context/comparison model of social influence and its leniency contract extension to the special case of in-group minorities suggest a pattern of members' interpersonal processes that unfolds over time to reconcile factions with the same social identity who hold different approaches to shared projects. Conditional on typical levels of task debate, we predict that (a) in early episodes, task conflict increases the relation between task debate and team performance; (b) in middle episodes, task conflict decreases the relation; and (c) in late episodes, task conflict increases the relation again. We explore our thesis using a longitudinal design with a sample of 60 student teams (360

Acknowledgments: The authors would like to thank Mike Zyphur, Fred Oswald, and three anonymous reviewers for comments and feedback that greatly improved our manuscript. In addition, the authors would like to thank Vanessa Monties for coding qualitative data and Mariëlle Zondervan-Zwijnenburg for counseling on complexity indices. This article builds on the first author's doctoral dissertation work at EMLYON Business School.

Corresponding author: Alan R. Johnson, EMLYON Business School, 23 Avenue Guy de Collongue, Écully 69130, France.

E-mail: johnson@em-lyon.com

individuals) working together for course credit over 5 months (21 weeks) to write a first business plan for a new venture. We use a multilevel structural equation modeling approach with Bayesian estimation. We found support for our theory expressed in informative hypotheses using Bayesian model selection. These results were not evident from conventional graphing and post hoc statistical probing of simple slopes against the null hypothesis.

Keywords: *in-group minority; interpersonal processes; debate and conflict; team performance; time; projects; alternative proposals*

From the late 1960s, organizations increasingly adopted team-based structures to respond to customers' demands (Salas, Goodwin, & Burke, 2008). Until then, top management assumed responsibility for most organizations' critical decisions. Today, project teams develop new products and services, self-managed work groups build and deliver them, and top management orchestrates activities from the boardroom. Where gaps in the market remain, new venture teams recombine resources to exploit those opportunities (Delmar & Shane, 2006; Klotz, Hmieleski, Bradley, & Busenitz, 2014). How teams selectively process information, adopt innovations, and commit to new goals are questions of critical importance. To address them, researchers have developed methods and theories to understand how team processes over time explain team performance (Marks, Mathieu, & Zaccaro, 2001). We offer an explanation of team performance based on a social influence perspective on interpersonal processes—the extent that people engage in social and psychological interactions with other members of their group, over time, which we ground in the minority influence strand of social influence theory (Crano, 2010; De Dreu & De Vries, 2001; Martin & Hewstone, 2008; Moscovici, 1985). This perspective allows us to suggest *why* members' task debate and task conflict over time *jointly* determine team performance, where task conflict moderates relations over time between task debate and performance.

Social influence theory conceptualizes processes of change in which people's behaviors (what they do or say) or cognitions (their attitudes, beliefs, and opinions), or both, are changed by the perceived cognitions and behaviors of others (Hogg, 2010). However, few studies have captured the richness of the temporal course of social influence by studying team performance as a function of interpersonal processes as they unfold over time. We conceptualize *two* dimensions of the interpersonal process construct (Marks et al., 2001): *task debate*, drawn from the small group discussion literature (e.g., Stasser & Titus, 2003), and *task conflict*, from the intragroup conflict literature (e.g., Jehn & Mannix, 2001), which reflect the extent to which members engage in *discussion* and *disagreement*, respectively. In previous team behavior literature, researchers argued for various dimensions of members' interpersonal processes to explain team performance because they allow people to share potentially useful cognitive resources, knowledge, and creativity with other members in their teams (e.g., M. A. West, 2002). If members do not engage with one another about their alternative proposals, their teams fail to access available resources and are less likely to succeed. However, unshared information does not readily emerge during face-to-face discussion in teams, and, when it does, it rarely is developed fully (Stasser & Titus). Furthermore, teams can be hotbeds of conflict that can stifle members' ideas, impose group conformity, and lapse into free riding (Jehn & Mannix).

Moscovici's (1976, 1985) insights into the social influence process suggest a framework where a dissenting *minority* prompts innovation, and ultimately increased team performance, through persistent, consistent, and unanimous advocacy (task debate). In response, at first, the majority tends to resist alternative proposals (*high* task conflict), but later, under some conditions, the majority may consider, or elaborate, the minority's proposals (*low* task conflict) to hear what they have in mind and, as a result, may validate it by adopting innovations related to a minority's original proposals. Crano and his colleagues (see Crano, 2010; Crano & Alvaro, 1998b, 2013; Crano & Seyranian, 2007, for reviews and citations) have extended Moscovici's theory in their context/comparison model (CCM) of social influence and social change and an extension labeled the leniency contract (LC). Their research and theory into the LC focuses on advocacy strategies that in-group minority factions may use to change majority members' attitudes over time. We extend LC from individual-level cognitions, beliefs, and attitudes to team-level behaviors and focus on in-group minorities who advocate proposals at odds with the position of the majority, in the context of small interacting task-oriented groups. We assume that intraindividual processes (i.e., how people think about arguments presented by an in-group minority) are reflected in interindividual behaviors (e.g., how members discuss minority arguments in their teams). Our research question is to explain why dissenting *in-group* minorities who swim against the majority tide and debate alternative proposals about their shared new venture creation projects induce conflict, then consensus, and then more conflict, which are catalysts for innovation that leads to improved team performance.

To test our theory in the fairest possible way, we use three novel methods. First, we use multilevel structural equation modeling (MSEM) because we need design, measurement, and analysis tools that allow us to specify the level *where* (team level) and the time *when* (early, middle, and late episodes) things happen (see Mathieu, Maynard, Rapp, & Gilson, 2008, for a review and citations). Second, we use Bayesian estimation (see Gelman, Carlin, Stern, & Rubin, 2004; van de Schoot, Kaplan, Denissen, Asendorpf, Neyer, & van Aken, in press; Zyphur & Oswald, 2015, for introductions) to overcome issues with maximum likelihood (ML) estimation. Third, we use informative hypotheses and Bayesian model selection to test our hypotheses directly (Hoijtink, 2012). The construction of informative hypotheses and the selection of the best fitting model with Bayes factors (BFs; Kass & Raftery, 1995) essentially involves a formal comparison of plausible scenarios suggested by scientific theory. More formally, an informative hypothesis is a testable proposition where background knowledge is expressed using inequality constraints that specify orderings for the parameter of interest. Methodologists from both classical and Bayesian model selection camps argue that comparison of two theoretically plausible scenarios (both translated into informative hypotheses) may be favored over comparing a theoretically plausible scenario with the null hypothesis (classical hypothesis testing; Harlow, Mulaik, & Steiger, 1997). Our approach to testing explicit expectations against one another, as informative hypothesis testing allows, is consistent with calls for theory refinement (Meehl, 1990) and more precise testing (Edwards & Berry, 2010) in the social and behavioral sciences, which builds on the classic debate on theory criticism in the philosophy of science (e.g., Lakatos & Musgrave, 1970). The key objection to the null hypothesis is that rarely, if ever, is it a plausible scenario based on scientific theory and past empirical research. According to Hoijtink, the best argument in favor of informative hypotheses is that on the basis of previous research, they allow researchers to evaluate their expectations directly against one another. We adopt procedures for this with MSEM analyses from van de Schoot and his colleagues (van de Schoot, Hoijtink, Hallquist, & Boelen, 2012; van de Schoot, Verhoeven, & Hoijtink, 2013).

The Leniency Contract

Hogg (2010) suggests the most widespread and enduring forms of social influence rest on conformity to self-relevant group norms, in which norms are configured by normative conflict within and between groups and among majorities, minorities, and factions. We adopt the LC model of minority influence to ground our hypotheses about team performance as a function of members' debate and conflict in small interacting groups. The LC is empirically validated in the laboratory and can be adapted to the field to conceptualize the richness of social influence processes unfolding over time between in-group minorities and majorities in interacting groups.

The LC is an extension of the CCM that addresses the special case of the in-group minority in social influence contexts. More generally, the CCM integrates features of social identity and self-categorization theory, inter- and intragroup relations theory, and classic persuasion theory to draw attention to several boundary conditions that have important effects on social influence processes but were not included systematically in the first 20 years of research on minority influence (Crano & Alvaro, 2013). The boundary conditions of the CCM include the targets' perceptions about the message, including (a) whether it emanates from a majority or minority source; (b) if it emanates from a minority source, whether it is from an in-group (i.e., an accepted faction of the larger group) or an out-group (i.e., *not* recognized as a legitimate part of the larger group); (c) whether it threatens the viability of the group, a source of members' social identity; (d) whether it admits to an objective (vs. subjective) judgment; and, finally, (e) whether it is supported by strong evidence or argumentation. The corresponding boundary conditions may apply to minority influence processes in new venture team contexts. Here, members have come together through homophily processes. Homophily refers to the selection of other members on the basis of similar characteristics, such as gender, ethnicity, nationality, appearance, and the like (Ruef, Aldrich, & Carter, 2003). Once affiliated with a team, members' social influence processes—task debate and task conflict—affect subsequent project choices that unfold over time by championing alternative proposals for their shared projects. The outcomes of these processes determine team performance.

The CCM distinguishes between minorities that are perceived as either an in-group or an out-group by the majority and holds that if a minority can establish and maintain in-group status, they enjoy an advantage in persuading the majority of the merits of their alternative proposals. The LC posits that in-group opinion minorities enjoy lenient consideration from majority members. While leniency from the majority does not result in immediate acceptance of an in-group minority's position, it attenuates counterargument of their message and inhibits derogation of their members. Over time, if an in-group minority has strong supporting evidence, they may introduce cognitive and behavioral conflict into the majority's belief system. Such a discrepancy can result in immediate change of the majority's attitudes that are related to the focus of debate (indirect change). If the indirect change is sufficient to unbalance the overall belief system, after further delay, this may prompt change of the majority's attitudes on the focal issue (direct change; Crano & Chen, 1998). In contrast, an out-group minority does not enjoy leniency when presenting a dissenting position and, as such, is unlikely to affect the majority's attitudes.

In early episodes, involving new venture teams, the majority may view in-group minorities who advocate alternative proposals (task debate) as disruptive and may even consider the minority as an out-group. Majority members may perceive a minority's proposals as generating unnecessary division and increased uncertainty within newly formed teams and, therefore, resist their influence attempts (*high* task conflict). In middle episodes, after the group is more established, it may consider alternate proposals to the generally accepted plan if provided by a minority that has proven to be an in-group and that has been persistent, consistent, and unanimous in its advocacy (task debate). In this circumstance, the majority may consider the alternate proposal open-mindedly and with little source derogation (*low* task conflict). Such elaboration by the majority may result in indirect change (i.e., change on related issues) consistent with the minority's petition. Finally, in late episodes, minority-induced indirect change that is sufficient to introduce cognitive and behavioral conflict towards the shared project may result in pressure towards consistency that rebalances the group's attitude structure (*high* task conflict). Thus, alternative proposals may prompt change on the focal (direct) issue addressed in the in-group minority's message, but only after some delay. These expectations are illustrated in Figure 1.

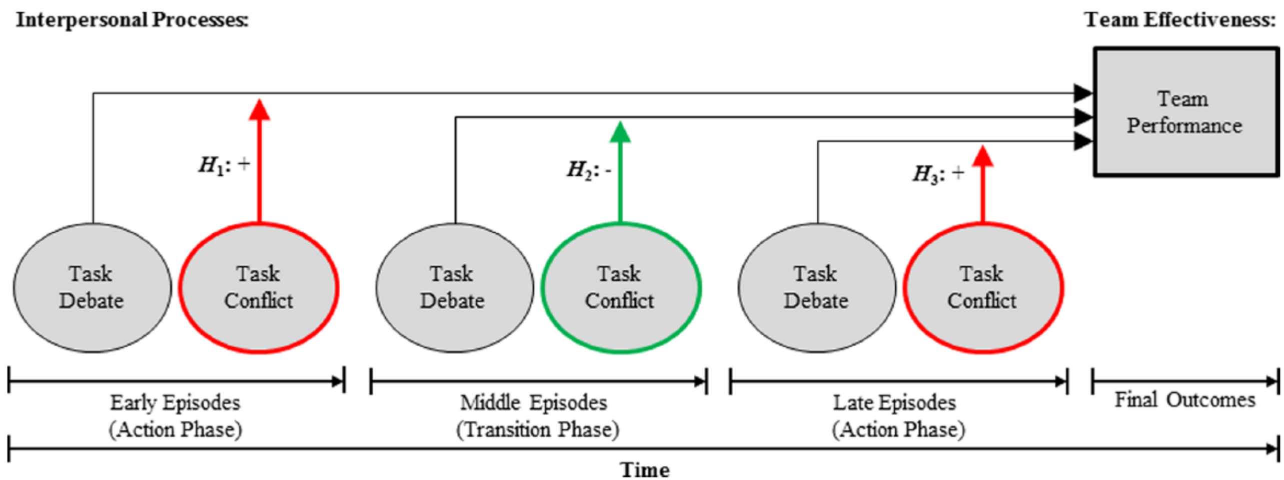
Using the model shown in Figure 1, we now specify our informative hypotheses as follows. Conditional on typical task debate across time, team performance with high task conflict will be *greater than* that with typical task conflict in early episodes (Hypothesis 1); in middle episodes, team performance with high task conflict will be *less than* that with typical task conflict (Hypothesis 2); and finally, in late episodes, team performance with high task conflict will be *greater than* that with typical task conflict (Hypothesis 3).[1] Testable propositions with inequality statements using *greater than* or *less than* are the sine qua non of informative hypotheses. As defined earlier, informative hypotheses include existing knowledge from the chain of evidence to propose testable propositions using inequality constraints about orderings of expected values for the parameter of interest—team performance—conditional plausible values for predictors, task debate and task conflict.

Influence Process Phases

Early Episodes

In early episodes, new venture teams strive to establish themselves, and, as such, their members are in a precarious position with respect to their social identities. Under these conditions, the majority may be quite intolerant of any alternative proposals, even from an in-group minority, as these may fractionate the group. Thus, despite the majority's unvested attitudes about their shared project, an in-group minority advocating an alternative proposal (task debate) may induce majority members' resistance as they cling to an uncertainty reducing orthodoxy with might and main (*high* task conflict). In comparison, *typical* task conflict is the ambient degree to which members disagree over the duration of their project, perhaps characterized by majority members who are not induced to resist because no in-group minority has advocated an alternative position.

Figure 1
Research Question and Conceptual Model



Note: H_1, H_2, and H_3 refer to Hypotheses 1, 2, and 3, respectively.

Our prediction in early episodes is grounded in the first assumption of the LC, which is based on the research result that despite persuasive arguments, in-group minorities typically do not induce direct (focal) attitude change in the majority (Wood, Lundgren, Ouellette, Busceme, & Blackstone, 1994). Crano and Alvaro (1998b) argue that while majority members may be reluctant to be identified with an in-group minority *position*, this implies neither that the majority strongly counterargues their proposals nor that they derogate, ostracize, or expel their members. Indeed, they have found that the majority's evaluation of in-group minorities improves as a function of the minority's presentation of arguments. The majority is unlikely to strongly counterargue or derogate an in-group minority precisely because they help define the social identity of all, and, thus, doing so poses a threat to their own social identity (Abrams & Hogg, 1990; Hogg & Abrams, 1988; Turner, 1991).

Considerable research on intra- and intergroup relations suggests that in-group members tolerate substantial heterogeneity for other in-group members' beliefs and attitudes. And, the out-group is assumed to be homogeneous ("they all think alike") and, typically, not in a good way (see Brewer & Brown, 1998, for a review). So, in-group variations in beliefs and attitudes between people are a potential and expected asset, not a danger to be avoided except on issues central to the group's identity (Worchel, Grossman, & Coutant, 1994). This may be the problem in early episodes, while members juggle their concerns about "getting on with the job" and "getting on with others" (Brown, 2000). People in task-oriented groups want to achieve some objective. To do this, they need to be tolerant of other in-group members' beliefs and attitudes. However, the soft stuff interferes with getting the job done, and so back to task-oriented activity. Thus, we hypothesize the following:

> *Hypothesis 1 (H_1):* With typical task debate among members during *early* episodes, we predict *greater* team performance with *high* task conflict than with *typical* task conflict.

Middle Episodes

In middle episodes, majority members are more likely to consider an in-group minority's alternative proposals for a shared project. The majority is motivated to understand *why* an in-group minority is advocating a dissenting position (debate), to resolve behavioral and cognitive conflict that an in-group minority induced in early episodes. Once the group is on a stable footing, an in-group minority proposal attracts majority members' consideration precisely because it challenges the prevailing norms in the now-established group. If the minority has established and maintained its in-group credentials, it can induce majority members to validate their proposal by carefully considering their arguments and engaging in intensive interaction with the advocates (*low* conflict), especially if the evidence is seen as based on objective data (vs. subjective judgments).

Our prediction in middle episodes is grounded in the second assumption of the LC, which is based on Petty and Cacioppo's (1986) dual-process model of persuasion, the elaboration likelihood model. In a persuasion context, *elaboration* is the extent to which people carefully consider the arguments relevant to the issue contained in a counterattitudinal message (Petty & Cacioppo). The model posits two modes of processing, central and peripheral. If people are able and motivated to process, they engage in central route processing to evaluate arguments and related information in a message. In this mode, message quality is the dominant determinant of message acceptance or rejection. When not able or motivated to process, people use peripheral cues (e.g., source status) to determine acceptance or rejection of the message.

Central route, high elaboration not only fosters attitude change but also affects attitude strength because when people make an effort to consider a message, they become invested in the outcome (Petty & Wegener, 1998). If people are neither willing nor able to think about the message and related information (i.e., *low* elaboration), their attitudes may be changed by peripheral processing, but these attitudes are weak, susceptible to reversals, and unlikely to guide behavior.

Given the lenient intragroup processes posited in the first assumption, by middle episodes, immediate change on the focal issue might be expected. However, the LC provides only a reasonable and courteous hearing of a minority's position, to maintain the viability and cohesion of the group. In exchange, a minority implicitly accepts that direct (focal) change is unlikely. Nevertheless, the LC does not imply that a minority is an impotent agent of change. The open-minded elaboration of an alternative proposal with little counterargument and no source derogation can create considerable change pressure. Although direct (focal) change to majority members' attitudes is precluded contractually, the pressure for change on them is real enough and appears in their related (indirect) attitudes. Furthermore, this change can be both insidious and profound because the majority does not resist change on attitudes they do not perceive to be in jeopardy (Alvaro & Crano, 1997). Thus, we hypothesize the following:

> *Hypothesis 2 (H_2):* With typical task debate among members during *middle* episodes, we predict *greater* team performance with *low* task conflict than with *typical* task conflict.

Late Episodes

In late episodes, if the previous phase produced a consensus, groups need to focus their energies on project completion. Thus, it is *now* functional for groups to exert conformity pressure towards the *new* consensus, to focus members on working toward newly established goals, and to guard against distraction by peripheral issues. This implies not only that it is hard for an in-group minority to make headway with dissenting proposals but also that laggards from the *old* majority who have not yet realigned their attitudes with the new order are under pressure to do so (*high* conflict). Consensus puts the group more or less in the opposite situation to where they were in early episodes. It now *must* be more entitative than it was in the beginning because the members now have a clear idea about *who* they are and *what* they are trying to accomplish (Yzerbyt & Demoulin, 2010). Under these circumstances, and with considerable pressure to "get on with the job," dissenting minorities are largely ignored and members from the former majority are denied the power of veto. All the members know that their group tenure is ending, so the LC's "get on with the others" motive is not as strong as it was.

Our prediction in late episodes is grounded in the third assumption of the LC that minority influence occurs by a minority articulating alternative proposals, which induce cognitive and social conflict in the majority. Over time, social influence occurs when majority members take steps to resolve the conflict by both questioning their own position and considering the minority's position (or one related to it) as a viable alternative (Martin & Hewstone, 2002). However, while majorities are typically motivated to maintain the status quo, a minority is empowered to advocate alternative proposals to attract, and perhaps demand, attention from the majority (Moscovici, 1976, 1985). Even so, minorities do well to remember that if they escalate to more coercive behavioral styles, then majorities retaliate, and they lose.

Crano and Alvaro (1998a, 1998b) point out that the "unrecognized quandary posed by this bargain" (1998a: 101) is that, over time, the majority is vulnerable on related (indirect) beliefs and attitudes associated with the focal issue of an in-group minority's alternative proposal. Moreover, after further delay, they may also be vulnerable on focal (direct) beliefs and attitudes in the proposal. This can occur because attitudes do not exist in a vacuum but, rather, are interrelated, to a greater or lesser extent, such that change of one component of the constellation of beliefs may affect other, related, components (see Eagly & Chaiken, 1998, for a review). Alternative proposals advocated by an in-group minority introduce stress to the system of interconnected attitudes that contains the focal belief. While the majority is typically motivated to maintain the status quo on the focal issue, they can defuse the stress by altering related (indirect) beliefs. So, after further delay, if there is sufficient disruption to the belief system, the focal issue may also move to restore balance with the surrounding related beliefs (Crano & Chen, 1998). However, delayed focal change is absent in those who show only a small change in their indirect attitudes. Thus, we hypothesize the following:

Hypothesis 3 (H_3): With typical task debate among members during *late* episodes, we predict *greater* team performance with *high* task conflict than with *typical* task conflict.

Putting the three assumptions together, the LC provides a context of relative tolerance for an in-group deviant faction to speak its piece freely and have its message processed with attention. Despite some initial resistance from majority members, the majority is expected to both (a) hear an in-group minority's message, without strong counterargument; and (b) *not* derogate their members, under ordinary conditions where they are not a threat to the group and, hence, majority members' identity as defined, in part, by their group membership. After some delay, elaboration of a strong message by the majority is likely, and with little or no counterargument or source derogation, conflict with the in-group minority is resolved and related (indirect) attitude change may occur. After further delay, conflict between beliefs resulting from the indirect change leads to direct (focal) attitude change on the issues addressed in the alternative proposal. Our interpretation of the LC as an integrated longitudinal theory of social influence in small group contexts allows us to put our three hypotheses together into a consolidated hypothesis:

Hypothesis 4 (H_4): H_1 and H_2 and H_3.

From a philosophy of science point of view, H_4 is much more specific than any of its component parts because there are more constraints to be fulfilled by the data, and the *riskier* the hypotheses, the more their theory is rewarded, if the hypotheses agree with the data (cf. Popper, 1963).

Method

Participants and Setting

Participants were 360 third-year students studying for a bachelor degree in business studies from a leading European business school. The course title was Programme de Création de Nouvelles Enterprises (PCE); it is mandatory for all students admitted to the bachelor program. The student teams wrote a first business plan for a nascent new venture. The school is one from the *grandes écoles* system of specialized universities in France that focus on offering a limited number of subjects to students (e.g., business, engineering, or aviation). These schools are prestigious and there is intense competition among students for admission. They have to follow *classe préparatoire* (special courses) offered in *Lycée* (regional colleges) and take (and do well in) exams before they can matriculate to a *grandes école*. The foregoing suggests that the students in the PCE are among the brightest in the country for their cohort. The students were 58% female (42% male) and were 20 years old on average ($SD = 0.65$, age range = 18–22 years). The students were nearly all French with just 7 students declaring other countries of origin.

In the first semester, students were randomly assigned to teams for a case-based course on new venture creation to acquaint them with the key concepts in entrepreneurship and the core components of a business plan. Between semesters, they self-selected themselves into 60 new teams (i.e., 6 members in each) for writing their business plan. The only restriction on team selection was that each had members from both genders. In the second semester, the student teams began work on their projects, which accounted for a large proportion of both their resources and their course grade from January to May 2007 (5 months). We reasoned that students' choices to join teams would be analogous to people's choices to commit their resources to start-up teams in the outside world. Homophily processes seemed to operate (Ruef et al., 2003); course records and face-sheet data from our surveys showed a clear relation between the team that students joined and the region where they made their *classe préparatoire*. Conversely, there was no relation apparent between team membership in the first and second semesters.

Protocol

There were nine professors (four full-time faculty and five teaching and research assistants) instructing on the PCE who led all workshops and heard all interim presentations during the program. An administrative assistant rotated the professors in a near random manner so that they worked with and assigned grades to different teams on different occasions. The professors provided interim feedback (both formal, i.e., interim grades, and informal, i.e., mentoring) to students on Weeks 1, 3, 6, 9, 13, 16, 17, 19, and 21 of the program. The professors maintained a file for each team containing data about the feedback they provided, the opportunity in their project, and members' attendance during the course.

Observer-rated performance. The first author recovered archival data on interim grades awarded to teams on nine occasions during the program. However, the teams' final grades were used for the course, audited by the course director, as the measure for cumulative team performance to Week 21. The final grading of each business plan was based on two professors' agreed appraisal of both an in vivo presentation by the team and a written document.

Alternative proposals for shared projects. Qualitative coding of team files by a research assistant suggests that in Week 1 (early episodes), only 5 teams had already selected the final proposal for their business plan and another 5 teams had narrowed the field down to two alternatives (17% of the total). By Week 9 (middle episodes), 13 teams had final proposals

Table 1
Individual-Level Correlation Coefficients, Means, and Standard Deviations From Ordinary Least Squares Estimation

Variable	1	2	3	4	5	6
1. Task Debate 1	—					
2. Task Debate 9	.36*	—				
3. Task Debate 17	.36*	.40*	—			
4. Task Conflict 1	–.25*	–.17*	–.31*	—		
5. Task Conflict 9	–.23*	–.36*	–.24*	.43*	—	
6. Task Conflict 17	–.19*	–.29*	–.37*	.35*	.52*	—
M	5.88	5.80	5.59	2.32	2.55	2.71
SD	0.82	0.88	0.89	0.89	1.00	1.06
Observations	346	342	216	346	343	217

*$p < .05$.

and another 24 had narrowed the field to two alternatives (62% of the total). By Week 17 (late episodes), 24 teams had final proposals and another 28 had just two alternatives (87% of the total). Examples of the opportunities exploited in proposals included selection and purchase of clothes for men, cookery courses from grandmothers, burial at sea, and so forth.

Missing data. The protocol for collecting student self-report data was that professors would bring surveys to class on the days when the teams were presenting their work in progress—Weeks 1, 9, 17, and 21—and each student would complete one at the beginning of class. However, 20 teams (33%) did not receive surveys in Week 17 because of an administrative error. We provide a summary of missing observations for both individuals and teams on each occasion in Tables 1 and 2, respectively. We argue that the data for 20 teams missing in Week 17 are "missing at random" (Van Buuren, 2012) because on each occasion, 10 teams were assigned to each professor in a classroom in a near random manner so that students saw projects from a variety of other teams and got feedback from a variety of professors. However, on Week 17, the student surveys were forgotten in two classrooms and, hence, the missing data for 20 teams. To compound the problem, these two classrooms were in a part of the campus buildings that, at that time, was unfamiliar to the first author. Our tale provides a rationale for our claim that the teams are missing at random in relation to our criterion variable. In addition, as a robustness check (e.g., McArdle, 2012), we estimated our model using complete case analyses (i.e., listwise deletion), and our results were qualitatively similar to the full information results that we report, but with slightly smaller BFs. Furthermore, we have nearly complete data (i.e., 59 out of 60 teams) for the predictors in Week 21; we ran the model including these measurements—a full information analysis with auxiliary variables—and our results were again qualitatively similar. Therefore, we conclude that the results presented are robust to different ways of treating missing data. We report results for full information below.[2]

Table 2
Team-Level (Cluster) Correlation Coefficients, Means, and Standard Deviations From Ordinary Least Squares Estimation

Variable	1	2	3	4	5	6	7
1. Task Debate 1	—						
2. Task Debate 9	.58*	—					
3. Task Debate 17	.54*	.51*	—				
4. Task Conflict 1	–.47*	–.42*	–.45*	—			
5. Task Conflict 9	–.33*	–.49*	–.31	.68*	—		
6. Task Conflict 17	–.44*	–.41*	–.64*	.57*	.68*	—	
7. Team Performance 21	.30*	.21	.31	–.28*	–.21	–.09	—
M	5.88	5.80	5.59	2.32	2.55	2.71	12.50
SD	0.43	0.46	0.56	0.52	0.62	0.61	3.17
Observations	60	60	40	60	60	40	60

*$p < .05$.

Team Measures

The students completed the identical 100-item self-report survey instrument on four occasions during the study—Weeks 1, 9, 17, and 21. We report here on students' responses in Weeks 1, 9, and 17 on predictor variables. There was a written guarantee of individual confidentiality printed at the beginning of each questionnaire. In addition, the first author made a short verbal presentation to all students about how individual responses would be confidential and any reporting would be untraceable to any particular person. For each team, more than 90% of members responded on each occasion (96% in Week 1; 95% in Week 9; 90% in Week 17).

Task debate. Our approach to measuring minority dissent differs from that used in laboratory experiments, where most social influence research has been conducted, because we tap into respondents' active advocacy of alternative proposals in the course of an ongoing social interaction, as opposed to their passive self-categorization based on feedback information about how they are compared to others. Our self-report questions were "We debate ideas on the work to be done," "We speak about alternative viewpoints on our tasks," and "We discuss several proposals for our work." The questions were designed to be answered on a 7-point scale anchored by 1 (*never*), 3 (*sometimes*), 5 (*often*), and 7 (*always*). We asked respondents to think about their experiences in their teams during the last several weeks when answering the questions.

In addition, our approach to measuring minority dissent also differs from that used in other field studies, which are much less numerous than experimental studies. De Dreu and West pioneered minority influence research in field settings and developed self-report questions to observe the extent that "a minority in a group publically opposed the beliefs, attitudes, ideas, procedures, or policies assumed by the majority" (2001: 1193; see also Kozlowski & Klein, 2000). Our approach to measuring minority dissent differs from De Dreu and West's because we focus on members' self-reports using "we" questions to reflect team-level minority advocacy (debate) of alternative proposals, or *making their case*, as opposed to the extent of opposition or disagreement with the majority. Indeed, this is also true of the intragroup conflict literature; while researchers do not typically cite the social influence literature, they suggest that members' task conflict and disagreement contains the agency for change in teams (e.g., Jehn & Mannix, 2001), as opposed to members' debates and discussions. We believe this may be a mistake, and while we hold these researchers in the highest esteem, we offer an alternative view. Behfar, Mannix, Peterson, and Trochim (2011) have also moved in this direction to focus more exclusively on members' debate and discussion; we differ from their position in that we retain the functional creation and resolution of conflict and disagreement.

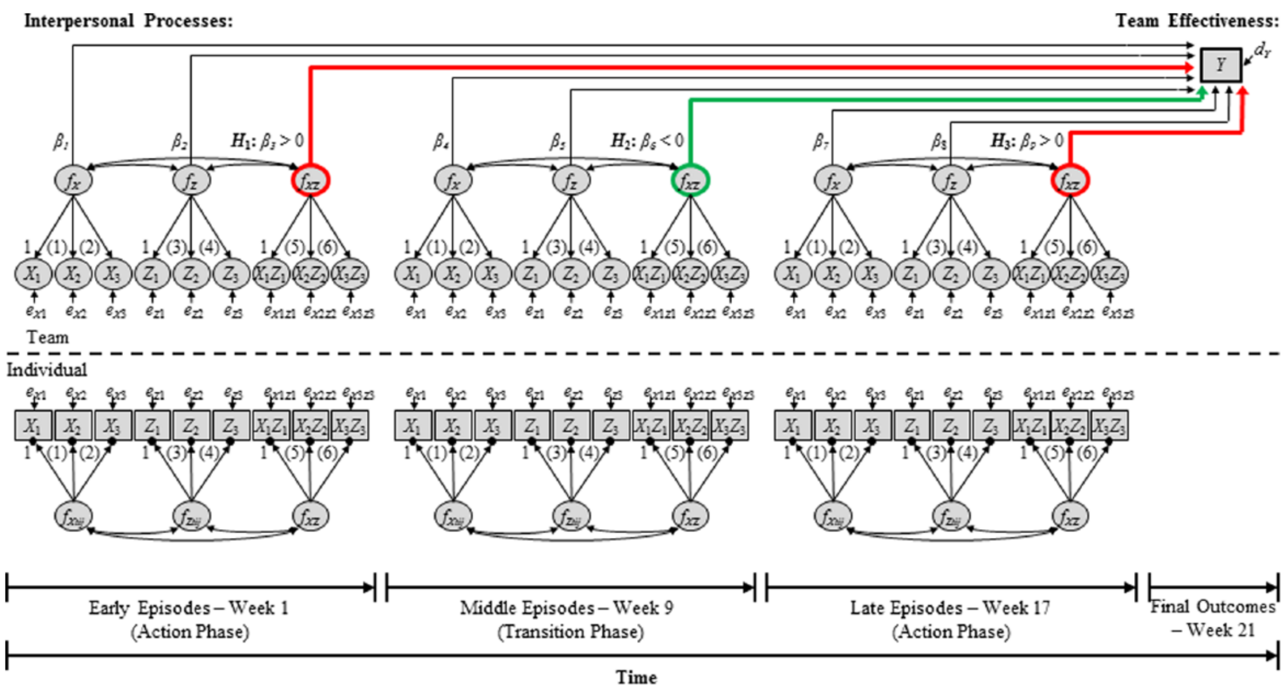
We present psychometric evidence supporting multilevel construct validity and invariance (Chen, Mathieu, & Bliese, 2004) for our three-question measurement of task debate on Weeks 1, 9, and 17. For each occasion, Cronbach's alphas were .89, .86, and .92, respectively, and the median values for the within-group agreement indices (r_{wg}; James, Demaree, & Wolf, 1984) were .94, .92, and .94, respectively, calculated using observed scores. Within-group reliability (ICC1) indices with 95% credibility intervals (CIs), which if the intervals do not include 0, show that the indices are significant at a $p < .05$ level, were .14, .21, and .35, with 95% CIs [.06, .27], [.11, .35], and [.20, .52], respectively, and reliability of group mean (ICC2) indices were .50, .62, and .77, with 95% CIs [.27, .69], [.42, .76], and [.60, .87], respectively, calculated using full information and Bayesian estimation.[3]

Finally, model fit indices exist for Bayesian estimation, for example, posterior predictive checking, but at the time of writing, posterior predictive checking is not implemented in *Mplus* (L. K. Muthén & Muthén, 2012) for MSEMs, so ML was used, where a selection of indices are available: chi-square values, the root mean square error of approximation (RMSEA), the comparative fit index (CFI), the Tucker-Lewis index (TLI), and standardized root mean square residual (SRMR). We note the cutoff values for the indices recommended by S. G. West, Taylor, and Wu (2012): RMSEA < .06, CFI and TLI > .95, and SRMR < .08.

We estimate univariate multilevel measurement models for task debate to demonstrate the measures have convergent validity, which means that the questions represent the same construct and their measurement is invariant over time. We report the following fit indices: $\chi^2(64) = 69.43$, $c = 0.72$, $p = .30$, RMSEA = .02, CFI = 1.00, TLI = 1.00, SRMR within = .03, SRMR between = .06, calculated from a three-wave single-factor MSEM where we restricted factor loadings and intercepts to be equal for each observed variable *both* across levels *and* over time, as noted in Figure 2. These fit indices are within the cutoffs. In addition, there was no significant reduction in fit between the full and restricted models from the chi-square difference test, $\Delta\chi^2(13) = 19.71$, $c = 0.96$, $p = .103$. Full results are available from the first author. We conclude that the measurement model is invariant, and, thus, we can interpret task debate consistently both at the team level and over time.

Task conflict. Our approach to measuring task conflict is directly analogous to that taken by intragroup conflict researchers (e.g., Jehn & Mannix, 2001). Our questions were "We have conflict over alternative ideas about our work," "There is discord in the team about viewpoints on our tasks," and "We disagree on our opinions about the work," anchored by the same 7-point scale described earlier.

Figure 2
Multilevel Structural Equation Model With Multiple Indicators and Repeated Measures



Note: The path diagram illustrates both the level *where* and the time *when* we posit things should happen in the model. We conceptualize our explanatory variables, represented by the f_Xs, f_Zs, and product terms f_{XZ}s at the team level, but we measure them at the individual level—represented by the *X*s, *Z*s, and product terms *XZ*s at the individual level. We both conceptualize and measure our outcome variable *Y* at the cluster level. We model the double-latent structure in the data to take care of both selection error and measurement error. We constrain some paths in the measurement parts of the model to equality in order to model invariance both across levels and over time. The boldface paths at the top of the figure represent the outcome variable *Y* regressed on the interaction terms f_{XZ}s that we refer to in our Hypotheses 1, 2, and 3 (H_1, H_2, and H_3, respectively).

We present comparable psychometric evidence for our three-question measurement of task conflict on Weeks 1, 9, and 17. For each occasion, we report Cronbach's alphas of .73, .77, and .81, respectively, and median values for within-group agreement indices (r_{wg}; James et al., 1984) of .92, .89, and .86, respectively, calculated using observed scores. We also report within-group reliability (ICC1) indices of .34, .40, and .30, with 95% CIs [.20, .49], [.26, .55], and [.14, .49], respectively, and group mean reliability (ICC2) indices of .75, .80, and .72, with 95% CIs [.60, .85], [.67, .88], and [.50, .85], respectively, calculated using full information and Bayesian estimation.

We also estimate univariate multilevel measurement models for task conflict to demonstrate convergent validity and measurement invariance over time. Model fit indices are $\chi^2(65) = 113.10$, $c = 0.80$, $p < .001$, RMSEA = .05, CFI = .95, TLI = .95, SRMR within = .05, SRMR between = .07, calculated from a three-wave single-factor MSEM with the same restrictions to factor loadings and intercepts as discussed earlier. The indices are again within the cutoff values. Again, there is no significant reduction in fit between the full and restricted models from a chi-square difference test, $\Delta\chi^2(16) = 18.42$, $c = 1.10$, $p = .300$, and we conclude that the measures are invariant.

We estimate a bivariate multilevel measurement model for both task debate and task conflict to demonstrate that the two sets of measures have discriminant validity, which means that they represent two different constructs. Model fit indices for the two-factor model are $\chi^2(272) = 389.69$, $c = 0.71$, $p < .001$, RMSEA = .04, CFI = .96, TLI = .96, SRMR within = .05, SRMR between = .10, calculated from a three-wave two-factor MSEM with the previously discussed restrictions. In comparison, the model fit indices for the one-factor model are $\chi^2(287) = 509.01$, $c = 0.70$, $p < .001$, RMSEA = .05, CFI = .94, TLI = .93, SRMR within = .09, SRMR between = .80. Comparing the indices for the two models, all are within sight of the recommended cutoff values (S. G. West et al., 2012), except for the SRMR between value in the one-factor model, which is 10 times more than recommended. The SRMR between results provide considerable evidence for discriminant validity between the two construct dimensions—task debate and task conflict—over time. However, taken together, the results also illustrate two important points about interpreting absolute model fit statistics in multilevel measurement models.[4] First, given our team-level thesis, the CFI and TLI are largely irrelevant because they are dominated by model fit in the individual part of the model (cf. Figure 2). The SRMR between is the only absolute model fit index that explicitly evaluates fit in the team part of the model. Second, methods scholars are still working on how precisely to interpret model fit indices in MSEMs to advise applied researchers on when a model does *not* fit. We like the severity of West and his colleagues' cutoff values, but we note that they are based on technical work with single-level models and may not *yet* be fully applicable to SRMR between indices. We acknowledge that exceeding the cutoff value for SRMR between places a limitation on our results, but we suspect that limitation is relatively minor. Additional evidence supporting discriminant validity is available from a chi-square difference test that also explicitly evaluates fit to the data in the team part of the model. The one-factor restricted solution significantly reduces model fit to the data, $\Delta\chi^2(15) = 119.32$, $c = 0.76$, $p < .001$, compared to the two-factor solution.

Descriptive Statistics

We provide descriptive statistics for both individual and team levels of analysis in Tables 1 and 2. At the team level, the correlations between predictors are generally high and significant, with task debate and task conflict negatively correlated. However, correlations between the outcome variable and predictors are only moderate in size, and occasionally significant, with task debate positively correlated with performance and task conflict negatively correlated with it. In addition, the sample mean values for task debate are generally declining during the study, whereas those for task conflict are increasing.

Treatment of Data

Why multilevel analysis? In recent decades, multilevel analysis has gained popularity in theoretical and applied research in the social and behavioral sciences because it can both describe and analyze relations between variables and constructs that are at different levels of analysis (Hox, 2010). With *micro-to-macro* models, researchers measure the outcome variable Y_j at the cluster level and explain it with predictor variables X_{ij}s measured at the individual level. For team behavior researchers, the conventional practice for analyzing data in this format, subject to some psychometric evidence criteria, is to aggregate individual-level measurements X_{ij}s to the cluster-level $\bar{X}_{.j}$s and then do regression analyses of the cluster-level outcome variable Y_j on the aggregated predictor variables $\bar{X}_{.j}$s. Croon and van Veldhoven (2007) have shown that conventional practice yields biased parameter estimates for regression coefficients (βs) because the practice fails to model sampling error in observed variables, and the bias in the estimates is proportional to the reliability of variables' cluster means, indexed by ICC2s (Lüdtke, Marsh, Robitzsch, Trautwein, Asparouhov, & Muthén, 2008). They have also shown that modeling sampling error using *best linear unbiased predictors* f_{Xj}s yields unbiased estimates of coefficients (βs). Applying Croon and van Veldhoven's critique to model both measurement and sampling error suggests a "doubly latent" MSEM approach to correct for both sources of random error (Lüdtke, Marsh, Robitzsch, & Trautwein, 2011). This critique provides the catalyst for our MSEM as shown in Figure 2.

Why Bayesian estimation? Despite the sophistication of the latent variable modeling approach in our MSEM, ML estimation may *not* detect substantively important effects (i.e., Type I error, failing to reject the null hypothesis). We have mentioned some limitations of ML; here, we outline several pragmatic ways researchers can use Bayesian statistics to overcome them and, in so doing, help detect important effects that otherwise might be missed. First, ML estimation assumes multivariate normality; Bayesian estimation does not. We model product term latent variables f_{XZ}s for our moderated model using product term indicators *XZ*s from observed variables *X*s and *Z*s (Marsh, Wen, Nagengast, & Hau, 2012). Product terms are known to have nonnormal distributions (Shrout & Bolger, 2002), and Bayesian estimation provides a solution (Yuan & MacKinnon, 2009). Second, with complicated models, ML estimates often fail to converge and, thus, provide unusable results (Singer & Willett, 2003). With uninformative priors, Bayesian estimation is more likely to converge and can provide results that are comparable with ML, had it converged (B. O. Muthén & Asparouhov, 2012). Third, with MSEMs and small sample sizes at the cluster level, ML parameter estimates are biased (too low) and standard errors are inefficient (too low; Hox, Maas, & Brinkhuis, 2010; Hox, van de Schoot, & Matthijsse, 2012). With the same data, Bayesian parameter estimates are about right, but standard errors are also inefficient (too high). Fourth, to model missing data with multiple imputations, ML not only uses multivariate normal distributions in the imputation process but also limits the number of imputed data sets to typically 5 or 10. Bayesian estimation uses empirically based multivariate distributions and allows the number of imputed data sets to equal the number of iterations in the Gibbs sampler—360,000 iterations for our final model—yielding more precision than ML (see Van Buuren, 2012). So, we use Bayesian estimation here, in a *pragmatic* way, which offers increased scope for estimating more complicated models (B. O. Muthén & Asparouhov).

Why Bayesian model selection? Besides pragmatic reasons for Bayesian statistics, testing the null hypothesis may *not* detect substantively important effects, even with Bayesian estimation (see Klugkist, Wesel, & Bullens, 2011; van de Schoot, Hoijtink, Mulder, Orobio de Castro, Meeus, & Romeijn, 2011; Wagenmakers, 2007, for discussions). Therefore, we use informative hypotheses to learn more from the data by evaluating support for plausible scenarios based on previous research and theory, versus one or more plausible alternatives. Thus, our informative hypotheses state inequality constraints that specify orderings for the parameter of interest, team performance Ys, which are expected values for $\hat{Y}$s that are conditional on plausible values of predictor variables, task debate Xs and task conflict Zs, based on previous research and theory from the field (Hoijtink, 2012: 5). Specifically, in early episodes, we hypothesize two conditional expected values for team performance $\hat{Y}$: first, team performance conditional on mean values for both predictors X and Z, labeled $\hat{Y}_1$; second, team performance conditional on mean values of X and values for Z that are 1 *SD above* the mean, labeled $\hat{Y}_2$. Thus, using these labels for conditional values of team performance to specify their expected orderings with inequality constraints, we state our first informative hypothesis, H_1: $\hat{Y}_2 > \hat{Y}_1$. Essentially, we are comparing the expected values of task performance $\hat{Y}$ under two sets of constraints, (a) *typical* task debate and *typical* task conflict and (b) *typical* task debate and *high* task conflict, and we expect $\hat{Y}_2$ *to be greater than* $\hat{Y}_1$ ($\hat{Y}_2 > \hat{Y}_1$). We used analogous labels for team performance to state our hypotheses for middle and late episodes, H_2: $\hat{Y}_4 < \hat{Y}_3$ and H_3: $\hat{Y}_6 > \hat{Y}_5$, respectively, where $\hat{Y}_3$ and $\hat{Y}_5$ are conditional on mean values of both predictors and $\hat{Y}_4$ and Y_6 are conditional on mean values of X and values for Z that are 1 *SD above* the mean.

Classical one-sided hypothesis testing can evaluate some specific hypotheses, in particular, a hypothesis with a single parameter and a sequence of hypotheses to compare one another where parameters are constrained versus unconstrained. However, three issues make this strategy difficult with our informative hypotheses. First, in our fourth hypothesis, H_4: ($\hat{Y}_2 > \hat{Y}_1$) & ($\hat{Y}_4 < \hat{Y}_3$) & ($\hat{Y}_6 > \hat{Y}_5$), we test more than one parameter, making one-sided testing difficult but not impossible. Second, while a "thoughtful frequentist approach" using planned comparisons overcomes the problem with multiple nonindependent tests (Type I error inflation), the test is still against the null hypothesis (van de Schoot, Hoijtink, et al., 2011). Even with classical model selection tools—Chi-square, Akaike information criterion, and Bayesian information criterion—the tests are still against the null hypothesis (the parameters are equal) versus the unconstrained alternative (the parameters are unconstrained). However, to test an informative hypothesis (a parameter is larger than another parameter), we recommend a direct test of two or more theory driven scenarios against each other. Third, we could impose inequality constraints on the parameter of interest, but then we could no longer use the classical tools to compare models (see Romeijn, van de Schoot, & Hoijtink, 2012).

Therefore, we formulate our informative hypotheses using inequality statements and evaluate them with BFs (Hoijtink, 2012; Klugkist, Laudy, & Hoijtink, 2005), where an informative hypothesis H_i with a BF of 2 is said to fit the data twice as well as the comparison model. To compute a BF for an informative hypothesis with MSEMs, we used the procedures described in van de Schoot et al. (2012), explained step-by-step in van de Schoot et al. (2013), and summarized in the appendix. Thus, we use BFs to choose the most plausible scenario (Hoijtink).

Analysis Strategy

Our presentation of results proceeds as follows: We establish moderated relations, probe the relations post hoc, and select the preferred model using BFs. We establish moderated relations by using an *F* test based on the R^2s for the criterion variable *Y*, team performance, provided in the *Mplus* (L. K. Muthén & Muthén, 2012) output with Bayesian estimation. This *F* test allows us to assess whether the *full* model, which includes product terms *XZ*s, explains *significantly* more variance ΔR^2 in the criterion variable *Y* compared to a *restricted* model, where the effects of the product terms are constrained to zero β_{XZ}s = 0. Then, we probe the results post hoc to interpret *significant* moderator effects by (a) statistical analyses of expected values for simple slopes considering significant differences from 0 (the null hypothesis) and (b) graphing the relations between *X* and *Y* at meaningful values for *Z* in each episode. Finally, we select the preferred model using BFs by analysis of the conditional expected values of the parameter of interest with inequality statements specified in the informative hypotheses. The first two steps are well understood by most team behavior researchers (cf. Aiken & West, 1991; Cohen & Cohen, 1983). However, the third step, using informative hypotheses and BFs, is relatively new and warrants further explanation, which we provided, partly, in the previous section and elaborate on more fully in the appendix.[5]

Results

Establishing Moderated Relations

We evaluated the *first-order effects* from performance *Y* regressed on both task debate *X*s and task conflict *Z*s on all occasions. Then, we evaluated the additional *second-order effects* from adding the product terms *XZ*s. We provide results from Bayesian estimation of our MSEM with the full information approach to missing data. The first-order effects explain a substantial portion of the variance (R^2 = 51.6%), but for each of our hypotheses, the model including the second-order effects explains significantly more variance ΔR^2s—H_1: ΔR^2 = 9.5%, $F(1, 52) = 12.70$, $p < .001$; H_2: ΔR^2 = 8.6%, $F(1, 52) = 11.24$, $p = .002$; H_3: ΔR^2 = 9.1%, $F(1, 52) = 12.04$, $p = .001$; H_4: ΔR^2 = 25.1%, $F(3, 50) = 17.95$, $p < .001$. These results support the hypothesized moderated relations but reveal nothing about the extent or direction of the change to the first-order relations. However, the significance of the overall test for a moderation effect justifies researchers probing the results in more detail. A conventional analysis aggregating predictor variables to the group level subject to some psychometric criteria (cf. Croon & van Veldhoven, 2007) would stop here because it does not yield any significant increases in variance explained ΔR^2 for any of our hypotheses. Further details of the conventional results are available from the first author.

Table 3
Results for Post Hoc Expected Values for Performance Y_{21} With z Scores From Bayesian Estimation

		Team performance						
		Mean task conflict (Z_M)			High task conflict ($Z_{M+1\,SD}$)		Low task conflict ($Z_{M-1\,SD}$)	
Expected value	Slope/ outcome	Estimate	*SE*	Slope/ outcome	Estimate	*SE*	Estimate	*SE*
Simple intercept								
Week 1		–0.36	344.50		–0.08	220.44	–0.64	599.02
Week 9		–0.39	272.51		–1.30	398.09	0.57	473.87
Week 17		–0.13	237.19		0.55	434.61	–0.68	477.59
Simple slope								
Week 1	[s_1]	0.75	797.98	[s_2]	0.94	829.33	0.57	785.54
Week 9	[s_3]	–0.17	247.25	[s_4]	–0.44	155.66	0.11	356.69
Week 17	[s_5]	0.14	82.38	[s_6]	0.18	176.45	0.14	198.49
Conditional outcome								
Week 1	[Y_1]	–0.21	146.85	[Y_2]	0.09	297.08	–0.53	352.79
Week 9	[Y_3]	–0.38	274.93	[Y_4]	–1.27	394.33	0.57	479.35
Week 17	[Y_5]	–0.19	233.24	[Y_6]	0.48	373.42	–0.76	490.29

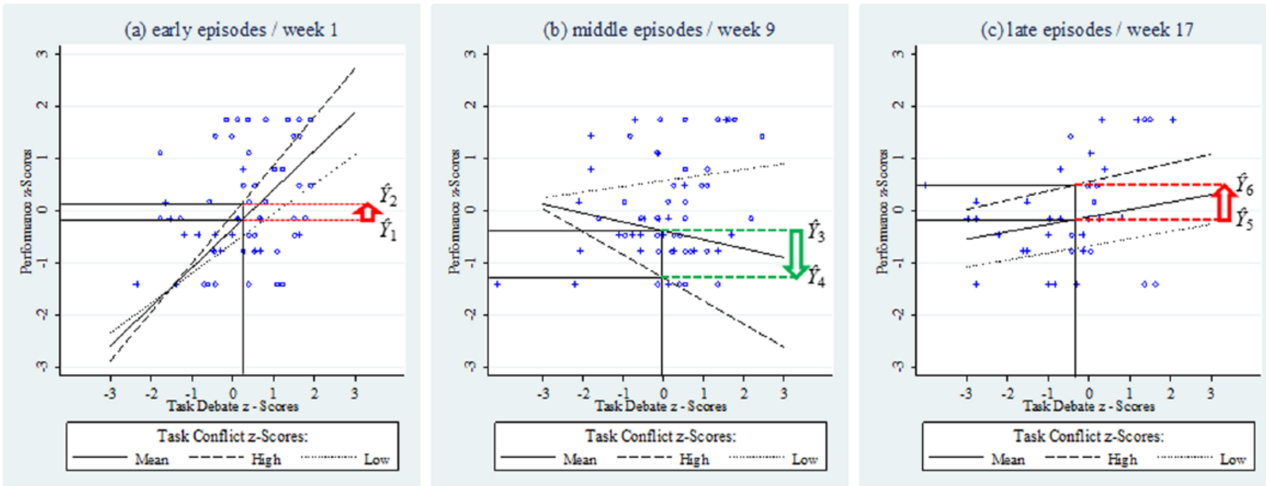
Note: Cluster-level sample size is 60.
*$p < .05$.

Post Hoc Probing

We proceed towards interpreting these estimates substantively by arranging them into simple regression equations to (a) probe conditional expected values post hoc and (b) graph them. Following the procedures for interpreting moderated relations outlined in Aiken and West (1991), we used the parameter estimates (βs) to calculate expected values for simple intercepts, simple slopes, and conditional outcomes. Table 3 tabulates the expected values from the simple regression equations and Figure 3 graphs the simple regression lines.

Table 3 presents the expected values from simple regression equations of Y regressed on X conditional on Z_M, $Z_{M+1\,SD}$, and $Z_{M-1\,SD}$ on each occasion (i.e., Weeks 1, 9, and 17) and controlling for the effects on other occasions. Aiken and West (1991: 14) suggest that researchers ask two questions: (a) For a specified value of Z, is the regression of Y on X significantly different from 0? and (b) For any pair of simple regression equations, do their slopes differ from one another? However, there are no significant simple slopes corresponding to our hypotheses. Thus, we do not find post hoc support from conventional comparison of simple slopes. In addition, the difference between conditional outcomes are not significantly different from 0 ($\hat{Y}_2 - \hat{Y}_1 = .331$, $SD = 291.196$, $p = .456$; $\hat{Y}_4 - \hat{Y}_3 = -.948$, $SD = 342.129$, $p = .392$; $\hat{Y}_6 - \hat{Y}_5 = .622$, $SD = 368.120$, $p = .413$). Even using MSEM, with ML estimation, a latent variable analysis of moderated relations against the null hypothesis would

Figure 3
Post Hoc Expected Values for Team Performance in *z* Scores From Bayesian Estimation



Note: Graphs illustrate z scores for (a) early, (b) middle, and (c) late episodes.

Table 4
Results From Bayesian Model Selection Using Bayesian Factors (BFs) With Performance $\hat{Y}$

	Team performance					
	Simple slope			Conditional outcome		
Comparison	f_i	c_i	BF	f_i	c_i	BF
Unconstrained (H_i vs. H_u)						
H_{i1}: $\hat{Y}_2 > \hat{Y}_1$	.54	.50	1.09	.54	.50	1.09
H_{i2}: $\hat{Y}_4 < \hat{Y}_3$	.60	.50	1.21	.61	.50	1.22
H_{i3}: $\hat{Y}_6 > \hat{Y}_5$	.50	.50	1.01	.59	.50	1.17
H_{i4}: $(\hat{Y}_2 > \hat{Y}_1)$ & $(\hat{Y}_4 < \hat{Y}_3)$ & $(\hat{Y}_6 > \hat{Y}_5)$	.19	.125	1.52	.34	.125	2.76
Alternative (H_{ia} vs. H_{ib})						
H_{i4} vs. H_{i1}			1.40			2.54
H_{i4} vs. H_{i2}			1.26			2.27
H_{i4} vs. H_{i3}			1.50			2.35
Complementary (H_i vs. $H_{\neg i}$)						
H_{i4} vs. $H_{\neg i4}$			1.64			3.68

Note: The sample size at the cluster level is 60 for the full information condition and 39 for the listwise deletion condition. Model fit indices f_i are estimated using 360,000 iterations. Model complexity indices c_i are calculated a priori to reflect the background probability of the model.

stop here because we found no significant differences between the conditional parameters of interest with this combination of data analysis tools.

However, Figure 3 reveals a pattern of relations between task debate X and performance Y conditional on *both* the extent of task conflict Z *and* the episode that is consistent with our informative hypotheses H_is. Figure 3a is consistent with our H_1, which illustrates with the *hollow upward* arrow that $\hat{Y}_2 > \hat{Y}_1$. Figure 3b is consistent with our H_2, which illustrates with the *hollow downward* arrow that $\hat{Y}_4 < \hat{Y}_3$. Figure 3c is consistent with our H_3, which illustrates with the *hollow upward* arrow that $\hat{Y}_6 > \hat{Y}_5$. Taken together, this pattern of results lends some support not only to our individual hypotheses H_1, H_2, and H_3 about the moderated effects of members' interpersonal processes over time on team performance but also to our LC theory about the pattern of these relations over time (H_4). Nevertheless, we need statistical evidence from post hoc probing of the functional relations to support this conclusion.

Model Selection

Table 4 displays the results from the Bayesian model selection analyses using BFs. Focusing on the results for the conditional outcomes $\hat{Y}$, these show that the single informative hypotheses H_1, H_2, and H_3 received *slightly* more support from the data than the unconstrained hypothesis H_u ($BF_{H1 \text{ vs. } Hu} = 1.09$; $BF_{H2 \text{ vs. } Hu} = 1.22$; $BF_{H3 \text{ vs. } Hu} = 1.17$). However, the consolidated informative hypothesis H_4 received more than *twice* the support from the data than the unconstrained hypothesis H_u ($BF_{H4 \text{ vs. } Hu} = 2.76$). In addition, the consolidated informative hypothesis H_4 received more than *twice* the support from the data than either of the other single informative hypotheses ($BF_{H4 \text{ vs. } H1} = 2.54$; $BF_{H4 \text{ vs. } H2} = 2.27$; $BF_{H4 \text{ vs. } H3} = 2.35$). Finally, the consolidated informative hypothesis H_{i4} received more than *three times* the support from the data than the complementary hypothesis $H_{\neg 4}$ ($BF_{H4 \text{ vs. } H\neg 4} = 3.68$). Comparing the BFs for conditional outcomes to those for simple slopes, there was about *double* the support from the data for the informative hypotheses H_is expressed in conditional outcomes. Taken together, these BF results lend moderately convincing support post hoc to the composite informative hypothesis H_{i4}, which is consistent with our LC theory of episodic social influence processes through members' interpersonal processes to determine team performance.

Discussion

In this study, we proposed that minority members can share their potentially new and useful ideas using interpersonal processes to have them adopted in a small group setting and, in so doing, transition to improved team outcomes via innovation. Specifically, we found that two dimensions of members' interpersonal processes—task debate (members' discussions) and task conflict (members' disagreements)—over time operate in opposition to one another to explain team performance. In early episodes, *above average* task conflict appears to *increase* the effect of *average* task debate on team performance. Conversely, in middle episodes, *below average* task conflict appears to *increase* the effect of *average* task debate on team performance. Finally, in late episodes, *above average* task conflict again appears to *increase* the effect of *average* task debate on team performance. However, the most important finding is that two dimensions of members' interpersonal processes operate *over time*, across early, to middle, to late episodes, to predict team performance in a relatively parsimonious and systematic way that we ground in social influence and social change theory (Asch, 1956; Moscovici, 1985).

Implications for Theory and Research

This study contributes to team behavior research and theory by conceptualizing members' interpersonal processes about the shared projects as two interacting dimensions of interpersonal interaction that determine team outcomes over time. Following social influence and social change theory, we posit that an opinion minority builds a case over time through their advocacy of dissenting proposals about shared group tasks. This process can take time for majority members to assimilate, even for in-group minority proposals (Crano & Chen, 1998), but the LC provides a framework to understand how the symbiotic relation between a minority's proposals and the majority's selective disposals, and sometimes adoptions, may play out over time for project teams from their formation to their task completion. The results suggest that a particular pattern of social influence processes can improve team performance in small groups episodically via subtle and alternating combinations of task debate and task conflict, where dissenting *in-group* minorities advocate alternative proposals about shared projects with others in the majority to induce behavioral and cognitive conflict, then consensus (i.e., resolve conflict), and then more conflict. We argue that these subtle and alternating combinations of interpersonal processes are catalysts for innovation, which lead to improved team performance. There are some counterintuitive elements to our theory, and yet as we try to broach in what follows, perhaps not as discrepant with previous theory as might appear at first blush. However, most of these apparent differences boil down to differences between a normative model of how teams should behave and a descriptive model of how they actually behave.

In early episodes, majority members generally disagreed with alternative views voiced by a minority, which tended to generate conflict within teams. Previous research and theory urges teams to audition several alternative proposals before deciding on the one to adopt, and equivalently, cautions against a premature consensus on the first proposal considered (De Dreu & West, 2001; Janis, 1972). One reason why potentially dissenting minorities go along with a premature consensus, without voicing their concerns, is that it is more comfortable to agree than to disagree, especially with no clear benefit from speaking up over time (Detert & Edmondson, 2011). On the basis of social identity theory and the LC, we have moved towards providing a road map to benefit potentially dissenting minorities. One reason business schools (including the one in our study) teach new venture teams "brainstorming" techniques (Paulus & Brown, 2003) is to encourage team members to produce and develop two or more alternative scenarios (Allen, 1977) and, thus, overcome their natural inclination to go along with others in their group (Asch, 1956). Simply put, if you are in a minority that plans to build an alternative scenario, then be prepared for a potentially hostile reception from the majority in your group, at least initially. However, if you believe that your alternative view is important for your team's success on the project, then persist in a reasonable and courteous way, and the others may come around to agree with you over time.

Later, in middle episodes, majority members may adopt a minority's dissenting point of view, but this usually takes some time and results in change on an (indirect) issue related to the one addressed in the original proposal. Still, the resulting consensus is often sufficient to resolve the conflict within the team and allow it to proceed with goal striving processes in a coordinated manner (cf. Gersick, 1988, 1989). We suggest that there are a couple of ways to read Gersick on midpoint transitions in her theory of group development. One is that a period of intensive interactions between members implies high task conflict, which is how Jehn and her colleagues (e.g., Jehn & Mannix, 2001) read Gersick before they commit to a final course of action. The other is that intensive interactions suggest task debate between members, which resolves previously accumulated task conflict that the in-group minority has induced by advocating discrepant data and leaves members free to commit to action. We believe that the second interpretation fits well with Gersick's notion for midpoint transitions because it came from her interpretation of Kuhn's notion of disrupted equilibrium in scientific progress (Gersick, 1992).

Alternatively, in middle episodes, majority members may not adopt a minority's dissenting point of view, perhaps because the proposal lacks sufficient merit to replace their current approach or perhaps because they doubt minority members' motives. De Dreu and Beersma (2001) acknowledge four motives for minority dissent in organizational settings—progressive, conservative, reactionary, and modernist. Only one of the four, progressive, characterizes a minority as advocates of an alternative proposal who seek to convince the majority of its value (Levine & Kaarbo, 2001). It is dangerous to assume that all dissenting voices are both meritorious and well meant because they may not be. This is why selective information processing, which is inherent in the three phases of our interpretation of the LC, is important. The majority can maintain group functioning if it receives a poorly argued proposal from a well-meaning in-group minority without adopting their proposal. A lenient hearing of their alternative views can be enough to allow many conflicts to be resolved, a consensus to be achieved, and goal striving processes to commence (Crano, 2001).

Finally, in late episodes, a minority's alternative views are likely to receive a similarly frosty reception as in early episodes but for different reasons. To encourage progress towards task completion, the majority now exerts consistency pressure on both any new minority and any laggards from the old majority to enforce the recently minted consensus achieved in middle episodes. In late episodes, there is time pressure to complete projects that tends to narrow the window for information processing available to consider any alternative points of view (Karau & Kelly, 2004). Taken together, the interplay between members' debate and conflict interpreted through social influence theory provides a compelling and temporally nuanced theory of how teams selectively adopt emergent creativity from their members without every dissenting faction holding team performance hostage.

The results contribute to team behavior research in design and methodology for studying team-level phenomena. We review the problems with aggregating individual-level responses of self-reported questions to the team-level predictor variables, for subsequent regression analysis on team-level criterion variables, because aggregating fails to model sampling error and, thus, leads to bias parameter estimates and inefficient standard errors (Croon & van Veldhoven, 2007). We show how a MSEM approach with Bayesian estimation not only can overcome aggregation biases by accounting for sampling error and measurement error (Lüdtke et al., 2011) but also can handle missing data modeling and convergence failures with several advantages compared to ML estimation (B. O. Muthén & Asparouhov, 2012). Finally, we adopt Bayesian model selection to choose our informative hypothesis with most support from the data (van de Schoot et al., 2012; van de Schoot et al., 2013) because of problems with classical hypothesis testing against the null hypothesis, even using Bayesian estimation. While we recommend this impressive set of tools, we acknowledge there remain both some links in the methodological chain that need strengthening and other legitimate ways to test hypotheses, which we return to in our Design Limitations and Future Research section. Nevertheless, it is important to highlight that Bayesian statistics, in general, and informative hypothesis testing and Bayesian model selection, in particular, are on the rise and likely to enrich future research throughout the social and behavioral sciences, and not just in team behavior research.

Managerial Implications

As teams increasingly take responsibility for critical business decisions, it is important for researchers and practitioners to understand how members' interpersonal processes determine team performance, in addition to inputs, emergent states, goal generation, and goal striving (Chen & Kanfer, 2006; Marks et al., 2001). Interpreted through social influence process theory, our study suggests a more prominent role for interpersonal processes than that suggested in previous research, which has emphasized the others more prominently. When doing new and complicated tasks, successful teams develop their ideas through a process by fostering team information exchange (Gong, Kim, Lee, & Zhu, 2013). Open exchange of information in collaborative efforts is critical for team effectiveness.

Members' interpersonal processes may increase the cognitive resources available within teams with a diverse member composition (M. A. West, 2002). However, differences between members tend to reduce the information they exchange (Mesmer-Magnus & DeChurch, 2009). Managers can introduce interim platforms or channels for exchanging ideas, perspectives, and knowledge, while members accommodate themselves to working across the fault lines in their teams (Lau & Murnighan, 1998). These procedures can help members voice their ideas (Paulus & Nijstad, 2003), which over time may foster a more supportive climate for innovation (Harrison, Price, Gavin, & Florey, 2002). However, there is a time for transition processes and a time for action processes, and if minority members are advocating transition while the majority members are engaged in action, this is likely to generate conflict (Hackman & Wageman, 2005).

Researchers from a functionalist perspective may be inclined to overestimate the informational component of interpersonal process determinants of team effectiveness at the expense of the normative component (Deutsch & Gerard, 1955). When invoked at the right time, both components have performance value for the team. The informational component is a vehicle for a persistent and courageous minority to have their alternative proposals considered before the group majority, in the hope of inducing adoption of some innovation. The normative component is a vehicle for members to coordinate themselves around some mature consensus and take action to complete their project. The team research and practitioner community may come to recognize the LC not only as a breakthrough cooperative strategy for managing intragroup conflict introduced by an in-group minority who advocates a counternormative alternative proposal but also as a necessary condition for maintaining group viability, which may not be explicit, or even conscious. In addition, the LC allows considerable latitude for in-group variation on many issues if members' deviance does not threaten the existence of the group and, thus, majority members' social identity.

Design Limitations and Future Research

We draw attention to some limitations to the study that should be borne in mind when interpreting its findings and implications. First, the study is longitudinal of team-level performance, but that does not rule out all alternative explanations to establish causality in the posited relations. Certainly, controlling for members' interpersonal processes at several theoretically relevant episodes during a project to explain team performance is important. Nevertheless, while the course archive contained team performance records for nine occasions, these were single-grade measures assessed by different professors. Thus, we are reluctant to assume invariance over time for the criterion variables and hesitate to use them in fixed effects regressions that would control for unobserved heterogeneity on any and all stable traits affecting the criterion variable (Allison, 2009). Thus, future research designs might consider multiple measures for criterion variables to test invariance and facilitate fixed effects analyses.

Second, we make some strong recommendations about how researchers should handle team-level predictors based on individual self-reports in micro-to-macro models. We base our recommendations on previous technical work in the methods literature and contrast findings there with prevailing practice in team behavior literature. However, in advocating the use of a variety of model fit indices (Kline, 2010), we find some fit anomalies in our data. While we believe these anomalies place only minor limitations on our findings, we urge further technical work on how applied researchers should interpret fit indices in the team part of the model (i.e., at the between-group level) from multilevel measurement models. Normal practice suggests that a poorly fitting measurement model precludes further analysis of the structural effects. However, we believe that our univariate measurement models demonstrate convergent validity that is temporally invariant, and they have absolute model fit indices within the cutoffs. We would argue that for the bivariate models, absolute fit is less important than the comparative fit for both the SRMR between index and chi-squared difference test because these metrics assess our claim that we have two constructs varying over time and not just one.

Third, this study is the first to examine members' interpersonal processes over time as determinants of team performance based on social influence processes within the episodic framework of team processes and outcomes. However, the episodic framework includes other team-level constructs, as listed in the last section. According to the framework, one or more of these might have mediated relations between members' interpersonal processes and team outcomes. Future research might profitably examine why and how the other team-level constructs mediate in some episodes but not in others.

Fourth, this study examined members' debate and conflict about alternative approaches to shared tasks as team processes that predicted team outcomes. However, the members involved in coalitions either advocating (debate) or resisting (conflict) alternative proposals from one occasion to the next may have varied. Theoretically, the LC suggests there is an advantage to having the same in-group opinion minority championing a consistent proposal over time. To investigate this, researchers would need to ask respondents to rate both the extent of team-level debate and conflict, as we did in this study, and the extent of each member's debate and conflict with each of the other members of the group. Specifically, respondents would answer both "*We* discuss alternative ideas on the work to be done" and "*Member X* discusses alternative ideas on the work to be done *with me*" for each member. Thus, the burden on respondents is considerably increased.

The above discussion leads to the more general question about how researchers should conceptualize and test bottom-up or micro-to-macro relations. Understanding these issues and the tools to implement them continues to progress in the research community. It is the researchers' responsibility not to allow a newfangled toolbox to beguile them with its "technobabble." Nevertheless, the MSEM and Bayesian tools from just several years ago are now maturing into tested and proven procedures that applied researchers can use confidently to produce robust findings. The key point for researchers to consider is *where* and *when* the phenomenon of interest is occurring and how best to measure it reliably. Overall, we hope this work will encourage further research into conceptualizing and testing micro-to-macro relations, which achieve improved understanding of team behavior phenomena.

Conclusion

We provide evidence that members' interpersonal processes over time—task debate and task conflict—combine in a systematic and parsimonious model to explain team performance in innovative and entrepreneurial contexts. We hope the study stimulates further research and theory into social influence and social change within the episodic framework for team behavior and outcomes. To facilitate future research and theory, we provide an instructive example of how to implement informative hypotheses and Bayesian model selection to social science research questions, which allow researchers to learn more from data than testing against the null hypothesis. These tools for articulating theory and fitting model to data have great promise for applied fields like team behavior, where data structures are nested, sample sizes at the cluster level are small, and observed variables often have skewed distributions. Finally, we submit that the LC has great promise for team behavior research and theory, especially as we move into more longitudinally nuanced theories of how and why individual actions and team processes unfold over time to determine team effectiveness.

Appendix

Bayes Factors

Researchers can use BFs (Kass & Raftery, 1995) to compare models in the Bayesian framework. Klugkist et al. (2005) and Hoijtink (2012) adopt the original BF for evaluating models with inequality constraints (see van de Schoot, Mulder, et al., 2011, for an introduction). An informative hypothesis is a testable proposition where background knowledge from the field is expressed using inequality constraints that specify orderings for the parameter of interest (Hoijtink). If there are contrasting accounts of a phenomenon in scientific discussion, then there are multiple informative hypotheses leading to several plausible orderings of the parameter of interest $\hat{Y}$s. To evaluate an informative hypothesis H_i, first compare orderings for the parameter of interest $\hat{Y}$s, suggested by background knowledge from the field, to an unconstrained hypothesis H_u that has no constraints imposed on any of the orderings. Researchers should make the comparison with H_u because it is possible that none of the informative hypotheses H_is under investigation provide an adequate description of the population. In that case, researchers should prefer the unconstrained hypothesis H_u, which prevents them from incorrectly choosing an inadequate informative hypothesis H_i.

The BF for comparing the informative hypothesis to the unconstrained hypothesis (H_i vs. H_u) is as follows:

$$BF_{H_i vs. H_u} = \frac{f_i}{c_i}, \tag{1}$$

where f_i is the model fit to the data and c_i is the a priori model complexity.

The model fit f_i is the proportion of the saved iterations from the Gibbs sampler where the order of the conditional parameters of interest $\hat{Y}$s from the informative hypotheses H_is agree with the data *in practice*. The Gibbs sampler is an iterative process where all parameters are freely estimated. If the inequality constraints for the conditional parameters of interest generally *agree* with the data, then the value for model fit f_i is *large*, for example, an estimate of .60 indicates that in 60% of the iterations from the Gibbs sampler, the estimated parameters agree with the constraints.

Model complexity c_i is the proportion of permutations where the order of pairs of conditional parameters of interest $\hat{Y}$s from the informative hypotheses H_is agree with the data *in theory* (Hoijtink, 2012: 49). If inequality constraints specify the order of *most* pairs of conditional parameters of interest $\hat{Y}$s, then the model is parsimonious and the model complexity c_i is *small*.

Tabulation of the permutations, highlighting those that are not allowed, and manually counting the remainder works fine for calculating complexity indices c_i with relatively small numbers of nonoverlapping constraints. However, where the numbers of constraints become large and start to overlap, the manual procedure becomes unwieldy. The second author has a doctoral student who has developed an R function to automate calculating complexity indices c_i.

In short, to evaluate support for informative hypotheses H_is, BFs are *large* to the extent that the model fit f_i is *large* or the model complexity c_i is *small* (see Hoijtink, 2012; van de Schoot, Hoijtink, et al., 2011, for more discussion). Researchers can interpret the BF ratio between model fit f_i and model complexity c_i in Equation 1 as the extent to which the data support their informative hypothesis H_i compared with the unconstrained hypothesis H_u. Roughly speaking, the comparison is between the *actual* rate that iterations from the Gibbs sampler fit with the informative hypothesis H_i compared to the *base rate* probability that the iterations will fit.

Returning to our study, the informative hypotheses H_1, H_2, H_3, and H_4 can be evaluated using Bayesian model selection. First, we compare these informative hypotheses H_is to the unconstrained hypothesis H_u. The unconstrained hypothesis H_u is always a BF = 1; therefore, BFs showing support from the data for informative hypotheses H_is need to be greater than 1. In addition, the BF for comparing the informative hypothesis to another informative hypothesis (e.g., H_1 vs. H_2) is as follows:

$$BF_{H_1 vs. H_2} = \frac{BF_{H_1 vs. H_u}}{BF_{H_2 vs. H_u}}, \tag{2}$$

where the BF for informative hypothesis H_1, calculated using Equation 1, is compared with the BF from H_2, also calculated using Equation 1. Finally, the BF for comparing an informative hypothesis with its complement (e.g., H_1 vs. $H_{\neg 1}$) is as follows:

$$\frac{BF_{H_1 vs. H_u}}{BF_{H_{\neg 1} vs. H_u}} = \frac{f_i / c_i}{(1-f)_i / (1-c_i)}, \tag{3}$$

where the BF for informative hypothesis H_1 in Equation 1 is compared with the BF from its complement informative hypothesis $H_{\neg 1}$. The complement of an informative hypothesis $H_{\neg 1}$ is interpreted as *not* H_1 (Hoijtink, 2012; van de Schoot et al., 2013). There are examples from our study of each BF in the Results section.

As Hoijtink (2012: 51) explains, there are no general guidelines for researchers to interpret BFs that are analogous to the benchmark of $p < .05$, which is typically used in frequentist statistics to interpret the significance. However, he suggests that most researchers agree that a BF = *1-point-something* is neither convincing evidence in favor of H_i nor convincing evidence against H_u. He suggests that researchers might interpret BFs > 1 as *slight* evidence in favor of the informative hypothesis H_i and BFs < 1 as *slight* evidence in favor of the unconstrained hypothesis H_u. Researchers should also note that with just one inequality constraint (i.e., H_1, H_2, and H_3), that the maximum value for a BF = 2, because the maximum model fit is always $f_i = 1$ and the complexity index $c_i = .5$ (i.e., $1 \div .5 = 2$). However, with three inequality constraints (e.g., H_4), the maximum value increases to $BF = 8$ and the maximum model fit remains at $f_i = 1$, but with the additional constraints, the complexity index reduces to $c_i = .125$ (i.e., $1 \div .125 = 8$).

We applied the four-step procedure as described by van de Schoot and his colleagues (van de Schoot et al., 2012; van de Schoot et al., 2013): First, use *Mplus* version 7.11 (L. K. Muthén & Muthén, 2012) with Bayesian estimation for the model, displayed in Figure 2, and save key post hoc expected values for conditional parameter estimates (i.e., simple slopes $\hat{s}$s, conditional outcomes $\hat{Y}$) all in one step. The result of the first step is a (quite large) file containing the conditional parameter estimates from each iteration of the Gibbs sampler. *Mplus* syntax for the final model is available from the first author on request. Two crucial elements when applying Bayesian statistics are (a) the priors used (we relied on the defaults in *Mplus*) and (b) convergence assessment. Concerning the latter, we used the Gelman-Rubin criterion (see Gelman et al., 2004, for more information) to monitor convergence, which is also a default setting in *Mplus*. Specify a minimum number of iterations using biterations=(90000) and request multiple chains for the Gibbs sampler using chains=8 and starting values based on the ML estimates using stvalues=ml. Finally, inspect all the trace plots manually to check whether all chains converged to the same target distribution and whether all iterations used for obtaining the posterior distribution were based on stable chains.

Second, for each informative hypothesis, calculate model fit f_i using an R package called *MplusAutomation* (Hallquist, 2012), which on the basis of the output from *Mplus* automatically counts the number of iterations where the order of the conditional parameters of interest $\hat{Y}$s from the informative hypotheses H_is *actually* agree with the data. Divide the number of "agreeing" iterations by the total number of iterations. For example, if the total number of iterations is 360,000 and the number of iterations where the parameters are in accordance with the informative hypothesis is 194,400, then f_{i1} = 194,400 ÷ 360,000 = .54.

Third, calculate the model complexity c_i manually by (a) counting the number of permutations where the order of the conditional parameter estimates $\hat{Y}$s in the informative hypotheses H_is agree with the data *in theory* and (b) divide by the total number of permutations. For example, where the number of permutations for conditional outcomes expected to agree *in theory* is 360, and the total number of permutations (6! = 6 × 5 × 4 × 3 × 2 × 1) is 720, then c_{i1} = .50 (i.e., = 360 ÷ 720).

Fourth, compute the BFs from Equations 1, 2, and 3.

Notes

1. High conflict is defined here as 1 *SD* above the mean, which is (a) standard in the team behavior literature and (b) recommended in the methods literature (Aiken & West, 1991; Cohen & Cohen, 1983).

2. On request, the first author can provide results from complete case analyses (i.e., listwise deletion) for comparison.

3. Very briefly, ICC1 is the proportion of the variance in a variable at the cluster level and is interpreted as within-group reliability—the correlation between individuals' self-reports on the variable in their team. ICC2 further divides the numerator of the ICC1 by the number of individuals per cluster and is interpreted as the reliability of the cluster mean for the variable.

4. We thank an anonymous reviewer for his/her patient prompting to sharpen this and other points in our Method section.

5. On request, the first author can provide the syntax we used with *Mplus* and *MplusAutomation*.

References

Abrams, D., & Hogg, M. A. 1990. Social identification, self-categorization and social influence. *European Review of Social Psychology*, 1: 195-228. doi:10.1080/14792779108401862

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

Allen, T. J. 1977. *Managing the flow of technology*. Cambridge, MA: MIT Press.

Allison, P. D. 2009. *Fixed effects regression models*. Thousand Oaks, CA: Sage.

Alvaro, E. M., & Crano, W. D. 1997. Indirect minority influence: Evidence for leniency in source evaluation and counter argumentation. *Journal of Personality and Social Psychology*, 72: 949-964.

Asch, S. E. 1956. Studies of independence and conformity: A minority of one against a unanimous majority. *Psychological Monographs: General and Applied*, 70: 1-70.

Behfar, K. J., Mannix, E. A., Peterson, R. S., & Trochim, W. M. 2011. Conflict in small groups: The meaning and consequences of process conflict. *Small Group Research*, 42: 127-176. doi:10.1177/1046496410389194

Brewer, M. B., & Brown, R. J. 1998. Intergroup relations. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed.), vol. 2: 554-594. New York: McGraw-Hill.

Brown, R. J. 2000. *Group processes: Dynamics within and between groups* (2nd ed.). Malden, MA: Blackwell.

Chen, G., & Kanfer, R. 2006. Toward a systems theory of motivated behavior in work teams. In B. M. Staw (Ed.), *Research in organizational behavior*, vol. 27: 223-267. Oxford, England: Elsevier.

Chen, G., Mathieu, J. E., & Bliese, P. D. 2004. A framework for conducting multi-level construct validation. In F. J. Yammarino & F. Dansereau (Eds.), *Research in multi-level issues*, vol. 3: 273-303. Oxford, England: Elsevier.

Cohen, J., & Cohen, P. 1983. *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Crano, W. D. 2001. Social influence, social identity, and group leniency. In C. K. W. De Dreu & N. K. De Vries (Eds.), *Group consensus and minority influence: Implications for innovation*: 122-143. Oxford, England: Blackwell.

Crano, W. D. 2010. Majority and minority influence in attitude formation and attitude change: Context/categorization—Leniency contract theory. In R. Martin & M. Hewstone (Eds.), *Minority influence and innovation: Antecedents, processes and consequences*: 53-77. New York: Psychology Press.

Crano, W. D., & Alvaro, E. M. 1998a. Indirect minority influence: The leniency contract revisited. *Group Processes & Intergroup Relations*, 1: 99-115. doi:10.1177/1368430298012001

Crano, W. D., & Alvaro, E. M. 1998b. The context/comparison model of social influence: Mechanisms, structure, and linkages that underlie indirect attitude change. *European Review of Social Psychology*, 8: 175-202. doi:10.1080/14792779643000119

Crano, W. D., & Alvaro, E. M. 2013. Social factors that affect the processing of minority-sourced persuasive communications. In J. P. Forgas, O. Vincze, & J. László (Eds.), *Social cognition and communication*, vol. 12: 467-497. New York: Psychology Press.

Crano, W. D., & Chen, X. 1998. The leniency contract and persistence of majority and minority influence. *Journal of Personality and Social Psychology*, 74: 1437-1450.

Crano, W. D., & Seyranian, V. 2007. Majority and minority influence. *Social and Personality Psychology Compass*, 1: 572-589. doi:10.1111/j.1751-9004.2007.00028.x

Croon, M. A., & van Veldhoven, M. J. P. M. 2007. Predicting group-level outcome variables from variables measured at the individual level: A latent variable multilevel model. *Psychological Methods*, 12: 45-57. doi:10.1037/1082-989X.12.1.45

De Dreu, C. K. W., & Beersma, B. 2001. Minority influence in organizations: Its origins and implications for learning and group performance. In C. K. W. De Dreu & N. K. De Vries (Eds.), *Group consensus and minority influence: Implications for innovation*: 258-283. Oxford, England: Blackwell.

De Dreu, C. K. W., & De Vries, N. K. (Eds.). 2001. *Group consensus and minority influence: Implications for innovation*. Oxford, England: Blackwell.

De Dreu, C. K. W., & West, M. A. 2001. Minority dissent and team innovation: The importance of participation in decision making. *Journal of Applied Psychology*, 86: 1191-1201. doi:10.1037/0021-9010.86.6.1191

Delmar, F., & Shane, S. 2006. Does experience matter? The effect of founding team experience on the survival and sales of newly founded ventures. *Strategic Organization*, 4: 215-247. doi:10.1177/1476127006066596

Detert, J. R., & Edmondson, A. C. 2011. Implicit voice theories: Taken-for-granted rules of self-censorship at work. *Academy of Management Journal*, 54: 461-488. doi:10.5465/AMJ.2011.61967925

Deutsch, M., & Gerard, H. B. 1955. A study of normative and informational social influences upon individual judgment. *Journal of Abnormal and Social Psychology*, 51: 629-636. doi:10.1037/h0046408

Eagly, A. H., & Chaiken, S. 1998. Attitude structure and function. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed.), vol. 1: 269-322. New York: McGraw-Hill.

Edwards, J. R., & Berry, J. W. 2010. The presence of something or the absence of nothing: Increasing theoretical precision in management research. *Organizational Research Methods*, 13: 668-689. doi:10.1177/1094428110380467

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. 2004. *Bayesian data analysis* (2nd ed.). Boca Raton, FL: Chapman & Hall.

Gersick, C. J. G. 1988. Time and transition in work teams: Toward a new model of group development. *Academy of Management Journal*, 31: 9-41. doi:10.2307/256496

Gersick, C. J. G. 1989. Marking time: Predictable transitions in task groups. *Academy of Management Journal*, 32: 274-309.

Gersick, C. J. G. 1992. Time and transition in my work on teams: Looking back on a new model of group development. In P. J. Frost & R. E. Stablein (Eds.), *Doing exemplary research*: 52-64. Newbury Park, CA: Sage.

Gong, Y., Kim, T.-Y., Lee, D.-R., & Zhu, J. 2013. A multilevel model of team goal orientation, information exchange, and creativity. *Academy of Management Journal*, 56: 827-851. doi:10.5465/amj.2011.0177

Hackman, J. R., & Wageman, R. 2005. A theory of team coaching. *Academy of Management Review*, 30: 269-287.

Hallquist, M. 2012. MplusAutomation: Automating Mplus model estimation and interpretation. An R package (Version 0.5-3). Retrieved from http://cran.r-project.org

Harlow, L. L., Mulaik, S. A., & Steiger, J. H. (Eds.). 1997. *What if there were no significance tests?* Mahwah, NJ: Erlbaum.

Harrison, D. A., Price, K. H., Gavin, J. H., & Florey, A. T. 2002. Time, teams, and task performance: Changing effects of surface- and deep-level diversity on group functioning. *Academy of Management Journal*, 45: 1029-1045. doi:10.2307/3069328

Hogg, M. A. 2010. Influence and leadership. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (5th ed.), vol. 2: 1166-1207. Hoboken, NJ: Wiley.

Hogg, M. A., & Abrams, D. 1988. *Social identifications: A social psychology of intergroup relations and group processes*. London: Routledge.

Hoijtink, H. 2012. *Informative hypotheses: Theory and practice for behavioral and social scientists*. Boca Raton, FL: CRC.

Hox, J. J. 2010. *Multilevel analysis: Techniques and applications* (2nd ed.). New York: Routledge.

Hox, J. J., Maas, C. J. M., & Brinkhuis, M. J. S. 2010. The effect of estimation method and sample size in multilevel structural equation modeling. *Statistica Neerlandica*, 64: 157-170. doi:10.1111/j.1467-9574.2009.00445.x

Hox, J. J., van de Schoot, R., & Matthijsse, S. 2012. How few countries will do? Comparative survey analysis from a Bayesian perspective. *Survey Research Methods*, 6: 87-93.

James, L. R., Demaree, R. G., & Wolf, G. 1984. Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology*, 69: 85-98.

Janis, I. L. 1972. *Victims of groupthink: A psychological study of foreign-policy decisions and fiascoes*. Oxford, England: Houghton Mifflin.

Jehn, K. A., & Mannix, E. A. 2001. The dynamic nature of conflict: A longitudinal study of intragroup conflict and group performance. *Academy of Management Journal*, 44: 238-251. doi:10.2307/3069453

Karau, S. J., & Kelly, J. R. 2004. Time pressure and team performance: An attentional focus integration. *Research on Managing Groups and Teams*, 6: 185-212. doi:10.1016/S1534-0856(03)06009-2

Kass, R. E., & Raftery, A. E. 1995. Bayes factors. *Journal of the American Statistical Association*, 90: 773-795. doi:10.2307/2291091

Kline, R. B. 2010. *Principles and practice of structural equation modeling* (3rd ed.). New York: Guilford Press.

Klotz, A. C., Hmieleski, K. M., Bradley, B. H., & Busenitz, L. W. 2014. New venture teams: A review of the literature and roadmap for future research. *Journal of Management*, 40: 226-255. doi:10.1177/0149206313493325

Klugkist, I., Laudy, O., & Hoijtink, H. 2005. Inequality constrained analysis of variance: A Bayesian approach. *Psychological Methods*, 10: 477-493. doi:10.1037/1082-989X.10.4.477

Klugkist, I., Wesel, F. van, & Bullens, J. 2011. Do we know what we test and do we test what we want to know? *International Journal of Behavioral Development*, 35: 550-560. doi:10.1177/0165025411425873

Kozlowski, S. W. J., & Klein, K. J. 2000. A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations*: 3-90. San Francisco: Jossey-Bass.

Lakatos, I., & Musgrave, A. (Eds.). 1970. *Criticism and the growth of knowledge*. Cambridge, England: Cambridge University Press.

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

Levine, J. M., & Kaarbo, J. 2001. Minority influence in political decision-making groups. In C. K. W. De Dreu & N. K. De Vries (Eds.), *Group consensus and minority influence: Implications for innovation*: 229-257. Oxford, England: Blackwell.

Lüdtke, O., Marsh, H. W., Robitzsch, A., & Trautwein, U. 2011. A 2 × 2 taxonomy of multilevel latent contextual models: Accuracy-bias trade-offs in full and partial error correction models. *Psychological Methods*, 16: 444-467. doi:10.1037/a0024376

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. 2008. The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods*, 13: 203-229. doi:10.1037/a0012869

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. 2001. A temporally based framework and taxonomy of team processes. *Academy of Management Review*, 26: 356-376. doi:10.5465/AMR.2001.4845785
Marsh, H. W., Wen, Z., Nagengast, B., & Hau, K.-T. 2012. Structural equation models of latent interaction. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling*: 436-458. New York: Guilford Press.
Martin, R., & Hewstone, M. 2002. Conformity and independence in groups: Majorities and minorities. In M. A. Hogg & R. S. Tindale (Eds.), *Blackwell handbooks of social psychology: Group processes*: 209-234. Malden, MA: Blackwell.
Martin, R., & Hewstone, M. 2008. Majority versus minority influence, message processing and attitude change: The source-context-elaboration model. In M. P. Zanna (Ed.), *Advances in experimental social psychology*, vol. 40: 237-326. New York: Academic Press.
Mathieu, J. E., Maynard, M. T., Rapp, T., & Gilson, L. 2008. Team effectiveness 1997-2007: A review of recent advancements and a glimpse into the future. *Journal of Management*, 34: 410-476. doi:10.1177/0149206308316061
McArdle, J. J. 2012. Foundational issues in the contemporary modeling of longitudinal trajectories. In B. Laursen, T. D. Little, & N. A. Card (Eds.), *Handbook of developmental research methods*: 385-410. New York: Guildford Press.
Meehl, P. E. 1990. Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry*, 1: 108-141. doi:10.1207/s15327965pli0102_1
Mesmer-Magnus, J. R., & DeChurch, L. A. 2009. Information sharing and team performance: A meta-analysis. *Journal of Applied Psychology*, 94: 535-546. doi:10.1037/a0013773
Moscovici, S. 1976. *Social influence and social change* (C. Sherrard & G. Heinz, Trans.). London: Academic Press.
Moscovici, S. 1985. Social influence and conformity. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (3rd ed.): 347-412. New York: Random House.
Muthén, B. O., & Asparouhov, T. 2012. Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, 17: 313-335. doi:10.1037/a0026802
Muthén, L. K., & Muthén, B. O. 2012. *Mplus user's guide* (7th ed.). Los Angeles: Muthén & Muthén. Retrieved from http://www.statmodel.com
Paulus, P. B., & Brown, V. R. 2003. Enhancing ideational creativity in groups: Lessons from research on brainstorming. In P. B. Paulus & B. A. Nijstad (Eds.), *Group creativity: Innovation through collaboration*: 110-136. New York: Oxford University Press.
Paulus, P. B., & Nijstad, B. A. 2003. An introduction. In P. B. Paulus & B. A. Nijstad (Eds.), *Group creativity: Innovation through collaboration*: 3-9. New York: Oxford University Press.
Petty, R. E., & Cacioppo, J. T. 1986. The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), *Advances in experimental social psychology*, vol. 19: 123-205. New York: Academic Press.
Petty, R. E., & Wegener, D. T. 1998. Attitude change: Multiple roles for persuasion variables. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed.), vol. 1: 323-390. New York: McGraw-Hill.
Popper, K. R. 1963. *Conjectures and refutations: The growth of scientific knowledge*. London: Routledge & Kegan Paul.
Romeijn, J.-W., van de Schoot, R., & Hoijtink, H. 2012. One size does not fit all: Derivation of a prior-adapted BIC. In D. Dieks, W. J. Gonzales, S. Hartmann, M. Stöltzner, & M. Weber (Eds.), *Probabilities, laws, and structures*: 87-106. Berlin: Springer
Ruef, M., Aldrich, H. E., & Carter, N. M. 2003. The structure of founding teams: Homophily, strong ties, and isolation among us entrepreneurs. *American Sociological Review*, 68: 195-222.
Salas, E., Goodwin, G. F., & Burke, C. S. (Eds.). 2008. *Team effectiveness in complex organizations: Cross-disciplinary perspectives and approaches*. New York: Routledge.
Shrout, P. E., & Bolger, N. 2002. Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, 7: 422-445.
Singer, J. D., & Willett, J. B. 2003. *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.
Stasser, G. L., & Titus, W. 2003. Hidden profiles: A brief history. *Psychological Inquiry*, 14: 304-313. doi:10.1080/1047840X.2003.9682897
Turner, J. C. 1991. *Social influence*. Buckingham, England: Open University.
Van Buuren, S. 2012. *Flexible imputation of missing data*. Boca Raton, FL: Taylor & Francis.

van de Schoot, R., Hoijtink, H., Hallquist, M. N., & Boelen, P. A. 2012. Bayesian evaluation of inequality-constrained hypotheses in SEM models using Mplus. *Structural Equation Modeling: A Multidisciplinary Journal*, 19: 593-609. doi:10.1080/10705511.2012.713267

van de Schoot, R., Hoijtink, H., Mulder, J., Orobio de Castro, B., Meeus, W., & Romeijn, J.-W. 2011. Evaluating expectations about negative emotional states of aggressive boys using Bayesian model selection. *Developmental Psychology*, 47: 203-212. doi:10.1037/a0020957

van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & van Aken, M. A. G. in press. A gentle introduction to Bayesian analysis: Applications to developmental research. *Child Development*. doi:10.1111/cdev.12169

van de Schoot, R., Mulder, J., Hoijtink, H., van Aken, M. A. G., Semon Dubas, J., Orobio de Castro, B., Meeus, W., & Romeijn, J.-W. 2011. An introduction to Bayesian model selection for evaluating informative hypotheses. *European Journal of Developmental Psychology*, 8: 713-729. doi:10.1080/17405629.2011.621799

van de Schoot, R., Verhoeven, M., & Hoijtink, H. 2013. Bayesian evaluation of informative hypotheses in SEM using Mplus: A black bear story. *European Journal of Developmental Psychology*, 10: 81-98. doi:10.1080/17405629.2012.732719

Wagenmakers, E.-J. 2007. A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review*, 14: 779-804. doi:10.3758/BF03194105

West, M. A. 2002. Sparkling fountains or stagnant ponds: An integrative model of creativity and innovation implementation in work groups. *Applied Psychology: An International Review*, 51: 355-387.

West, S. G., Taylor, A. B., & Wu, W. 2012. Model fit and model selection in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling*: 209-231. New York: Guilford Press.

Wood, W., Lundgren, S., Ouellette, J. A., Busceme, S., & Blackstone, T. 1994. Minority influence: A meta-analytic review of social influence processes. *Psychological Bulletin*, 115: 323-345. doi:10.1037/0033-2909.115.3.323

Worchel, S., Grossman, M., & Coutant, D. 1994. Minority influence in the group context: How group factors affect when the minority will be influential. In S. Moscovici, A. M. Faina, & A. Maass (Eds.), *Minority influence*: 97-114. Chicago: Nelson-Hall.

Yuan, Y., & MacKinnon, D. P. 2009. Bayesian mediation analysis. *Psychological Methods*, 14: 301-322. doi:10.1037/a0016972

Yzerbyt, V., & Demoulin, S. 2010. Intergroup relations. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (5th ed.): 1024-1083. Hoboken, NJ: Wiley.

Zyphur, M. J., & Oswald, F. L. 2015. Bayesian estimation and inference: A user's guide. *Journal of Management*, 41: 390-420. doi:10.1177/0149206313501200