



# Group Dynamics: Theory, Research, and Practice

Editor: David Marcus, PhD

[Journal Information](#)

[Current issue feed](#)

[Set PsycALERT: New issue update by email](#)



## 2014 Volume 18, Issue 2 (Jun)

### [Teamwork Competency Test \(TWCT\): A step forward on measuring teamwork competencies.](#)

Pages 101-121

Aguado, David; Rico, Ramón; Sánchez-Manzanares, Miriam; Salas, Eduardo

[Abstract](#) | [Full Text PDF](#)  | [Full Text HTML](#)  | [Permissions](#) | [Find It](#) 

### [The effect of different phases of synchrony on pain threshold.](#)

Pages 122-128

Sullivan, Philip J.; Rickers, Kate; Gammage, Kimberley L.

[Abstract](#) | [Full Text PDF](#)  | [Full Text HTML](#)  | [Permissions](#) | [Find It](#) 

### [Reactions to social inclusion and ostracism as a function of perceived in-group similarity.](#)

Pages 129-137

Sacco, Donald F.; Bernstein, Michael J.; Young, Steven G.; Hugenberg, Kurt

[Abstract](#) | [Full Text PDF](#)  | [Full Text HTML](#)  | [Permissions](#) | [Find It](#) 

### [Single-case designs in group work: Past applications, future directions.](#)

Pages 138-158

Macgowan, Mark J.; Wong, Stephen E.

[Abstract](#) | [Full Text PDF](#)  | [Full Text HTML](#)  | [Permissions](#) | [Find It](#) 

### [Team task conflict resolution: An examination of its linkages to team personality composition and team effectiveness outcomes.](#)

Pages 159-173

O'Neill, Thomas A.; Allen, Natalie J.

[Abstract](#) | [Full Text PDF](#)  | [Full Text HTML](#)  | [Permissions](#) | [Find It](#) 

### [The impact of group norms and behavioral congruence on the internalization of an illegal downloading behavior.](#)

Pages 174-188

Sansfaçon, Sophie; Amiot, Catherine E.

[Abstract](#) | [Full Text PDF](#)  | [Full Text HTML](#)  | [Permissions](#) | [Find It](#) 

## Teamwork Competency Test (TWCT): A Step Forward on Measuring Teamwork Competencies

David Aguado

Autonomous University of Madrid and Knowledge Engineering Institute

Ramón Rico

Autonomous University of Madrid

Miriam Sánchez-Manzanares

Carlos III University of Madrid

Eduardo Salas

University of Central Florida

The Teamwork Knowledge, Skill, Ability Test (TWKSAT), designed to assess teamwork competencies, has been widely used in both applied and academic contexts. However, studies have brought to light a number of reliability problems in the test. In this article, we describe 3 studies that (a) examined the functioning of the TWKSAT ( $N = 135$ ); (b) investigated a new measure, the Teamwork Competency Test, including its metric properties ( $N = 120$ ); and (c) analyzed the convergent and predictive validity of the Teamwork Competency Test compared with the TWKSAT ( $N = 91$ ). Based on our results, we conclude that the TWKSAT does not adequately reflect the initial substantive model and has limitations with regard to reliability. The Teamwork Competency Test improves the TWKSAT by enhancing reliability, content validity, and substantiating the dimensional structure of the test.

*Keywords:* teamwork test, teamwork competencies, team performance

Work teams have responded effectively to the global challenges of our times, providing organizations with the benefits of adaptability, productivity, and creativity above and beyond the contributions individuals can make on their own (Kozlowski & Ilgen, 2006; Salas, Sims, & Burke, 2005). However, teamwork requires several competencies to allow its members to effectively integrate their contributions, function as a unified whole, and make an “expert team”

out of a mere “group of experts” (Lawler & Worley, 2006).

Work teams bring to their mission much more than just the behaviors directly related to the task at hand. Their members need to interact and cooperate if they are to synchronize effectively (Salas et al., 2005; Salas, Sims, & Klein, 2004), which requires a specific set of behaviors that further the attainment of team goals. Empirical evidence shows that measures of knowledge and skills related to both a specific task and teamwork predict individual performance in work contexts (e.g., McClough & Rogelberg, 2003; Schmidt & Hunter, 1983). At the same time, teams formed by people with strong teamwork competencies display a specific range of behaviors, including the use of integrative (win–win) as opposed to distributive (win–lose) negotiating strategies. These findings are relevant as teams are often affected by interpersonal conflicts and faulty cooperation, which hinder optimum performance (Hackman, 2002). The availability of an appropriate measure of teamwork competencies may help minimize these problems (Stevens & Campion, 1994, 1999).

---

This article was published Online First May 12, 2014.

David Aguado, Autonomous University of Madrid, Madrid and Knowledge Engineering Institute; Ramón Rico, Autonomous University of Madrid; Miriam Sánchez-Manzanares, Carlos III University of Madrid; Eduardo Salas, University of Central Florida.

This research was partially supported by the Office of Naval Research Collaboration and Knowledge Interoperability Program and Office of Naval Research Multidisciplinary University Research Initiative Grant N000140610446.

Correspondence concerning this article should be addressed to David Aguado, Autonomous University of Madrid, Ciudad Universitaria de Cantoblanco, Madrid 28049 Spain. E-mail: david.aguado@uam.es

Over recent decades, scholars increasingly have underlined the importance of teamwork competencies for the effectiveness of groups and work teams (e.g., Cannon-Bowers & Bowers, 2011; Cannon-Bowers, Tannenbaum, Salas, & Volpe, 1995; Chen, Donahue, & Klimoski, 2004; Leach, Wall, Rogelberg, & Jackson, 2005; Peeters, van Tuijl, Rutte, & Reymen, 2006). In particular, Stevens and Campion (1994, 1999) proposed the Teamwork Knowledge, Skill, Ability Test (TWKSAT), which identifies and measures five transportable teamwork competencies that are common to different types of teams (see Figure 1) and have been related to different performance criteria (Mc-

Clough & Rogelberg, 2003; Stevens & Campion, 1999).

However, studies employing the TWKSAT have reported low levels of reliability (Athanasaw, 2003; Chen et al., 2004; McClough & Rogelberg, 2003). Despite the TWKSAT's capacity to predict individual performance in team tasks, there are no studies addressing the causes of the test's low reliability or explaining its dimensional structure. To resolve the problems affecting the assessment of teamwork competencies, we (a) examined the metric properties of the TWKSAT items, as well as their content validity and dimensional structure, and (b) propose a new measure based on the Stevens and

<b>I. Interpersonal KSAs</b>
<b>1. Conflict Resolution</b>
Recognize team conflict.
Recognize type and source of conflict and implement conflict resolution strategies.
Employ integrative (win-win) negotiation strategies.
<b>2. Collaborative Problem Solving</b>
Identify situations requiring participative group problem solving.
Recognize the obstacles to collaborative group problem solving and implement corrective actions.
<b>3. Communication</b>
Understand communication networks and utilize decentralized networks to enhance communication.
Communicate open and supportively.
Listen in a non-evaluative way and use active listening techniques.
Maximize consonance between non-verbal and verbal messages, and recognize and interpret the non-verbal messages of others.
Engage in ritual greetings and small talk.
<b>II. Self-management KSAs</b>
<b>4. Goal Setting and Performance Management</b>
Establish specific, challenging and accepted team goals.
Monitor, evaluate, and provide feedback on both overall team performance and individual performance.
<b>5. Planning and Task Coordination</b>
Coordinate and synchronize activities, information, and task interdependences.
Establish task and role expectations of individual team members, and ensure proper balancing of workload in the team.

Figure 1. The Stevens and Campion model.

Campion (1994, 1999) model, analyzing its reliability and validity as a measure of teamwork competencies. Although recently some tests have been proposed to capture knowledge on teamwork (e.g., boundary-spanning activities, Marrone, Tesluk, & Carson, 2007; team roles, Mumford, Van Iddekinge, Morgeson, & Campion, 2008), it is still necessary to develop tests capable of assessing teamwork competencies to predict employees' performance in teams and thus design and develop high-performance teams. We aimed to improve the assessment of teamwork-related competencies in work settings, beyond the TWKSAT, one of the most popular measures in the field.

### The TWKSAT as a Measure of Teamwork Competency

Competencies are defined as the underlying characteristics integrated with an individual's knowledge, skills, and abilities that are causally related to a referential criterion of effective and/or superior action in a specific job or situation (Spencer & Spencer, 1993). In particular, researchers have stressed the key role of teamwork competencies for the effectiveness of work teams (e.g., Cannon-Bowers et al., 1995; Chen et al., 2004).

Cannon-Bowers et al. (1995) identified eight central teamwork competencies, namely adaptability, shared understanding of the situation, performance monitoring and feedback, leadership, interpersonal relations, coordination, communication, and decision making. These scholars also made a key contribution by defining competencies on the basis of two dimensions: task competencies versus team competencies and specific competencies versus general competencies. Given their relationship with performance in different teamwork situations, general team competencies (i.e., transportable competencies) are of particular interest to organizations. With this in mind, Stevens and Campion (1994, 1999) identified five transportable teamwork competencies that are common to different types of teams: (1) conflict resolution, (2) collaborative problem solving, (3) communication, (4) goal setting and performance management, and (5) planning and task coordination. These competencies are grouped in two more general dimensions termed *interpersonal competencies* (defined by the first three items in the

list) and *self-management competencies* (represented by the last two items). This distinction is based on the idea that team effectiveness depends on its members' ability to both manage their relations appropriately and direct their actions to carry out the tasks assigned by the organization.

Perhaps the most important contribution made by Stevens and Campion (1999) is the construction of the TWKSAT, which provides both practitioners and academics with an instrument to measure key teamwork competencies. The available evidence suggests that the TWKSAT has considerable predictive validity. Thus, the TWKSAT measure of employees in real work teams correlated with their performance in the team as evaluated by both supervisors (correlations of between .23 and .52) and colleagues (correlations of between .21 and .34; McClough & Rogelberg, 2003; Stevens & Campion, 1999). Furthermore, Chen et al. (2004) found that the TWKSAT was sensitive to changes in the individual competency of university students after participation in a training program designed to develop their teamwork competencies.

However, other studies in which the TWKSAT was used have also consistently reported low reliability of the measure. Stevens and Campion (1999) originally found reliability of .80 (internal consistency), but this appears to be an overestimate (McClough & Rogelberg, 2003). In fact, the alpha coefficients in studies employing the TWKSAT were consistently lower. For example, McClough and Rogelberg (2003) found an alpha coefficient of .59, and Chen et al. (2004) found a coefficient of .64 before training in teamwork skills and .82 after training. Athanasaw (2003) obtained coefficients of .66 for the complete scale and between .25 and .48 for each of the five competencies, and Leach et al. (2005) found a coefficient of .70. As McClough and Rogelberg point out, however, the TWKSAT was designed from a multidimensional standpoint and other situational judgment tests share the low alpha coefficients found to date. In this type of test, internal consistency-based reliability measures should be complemented with test-retest estimations (Clevenger, Pereira, Wiechmann, Schmitt, & Harvey, 2001; McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001). However, Chen et al. reported a test-retest cor-

relation of .60 in the control group for their study. Overall, the available evidence suggests that the reliability of the TWKSAT could be improved, despite the test's predictive capacity.

No studies have been carried out to date to identify the reasons for the TWKSAT's low reliability, either by examining the metric characteristics of the items or the degree to which the different contents proposed by Stevens and Campion (1994) are actually present in the TWKSAT. Nor has any attempt been made to explain the dimensional structure of the TWKSAT, although this is crucial to identifying the scoring obtained from the test.

Overall, our main objective in this study was to present a new measure of the teamwork-related competencies based on Stevens and Campion's (1994) original model. To this end, we conducted three empirical studies. The first was devoted to analyzing the metric properties of the items and the content validity and dimensionality of the TWKSAT. In the second study, we proposed a new measure, testing the improvements made on reliability and dimensionality in a sample of university students. Finally, in the third study, we tested the validity of the new measure in a sample of professional employees.

### Study 1

The aim of this first study was to analyze the items composing the TWKSAT, examining their content and dimensionality. To begin, we present the results of the descriptive analysis based on the impact of each item on reliability. Following the strategy employed by Rovinelli and Hambleton (1977), we then go on to analyze the test items in terms of content validity. Next, the dimensionality of the test is examined using factor analysis.

### Method

**TWKSAT translation.** The TWKSAT comprises 35 multiple-response items, which describe different situations that may arise within a work team.<sup>1</sup> Respondents answer the items by indicating how they would act in each situation. The questionnaire evaluates the five teamwork competencies identified by Stevens and Campion (1994). An example item is the following (Stevens & Campion, 1999):

Your team wants to improve the quality and flow of the conversations among its members. Your team should: (A) use comments that build upon and connect to what others have already said; (B) set up a specific order for everyone to speak and then follow it; (C) let team members with more to say determine the direction and topic of conversation; (D) do all of the above.

The original version of the TWKSAT is written in English. We therefore applied the back-translation procedure (Brislin, 1970) to use the scale in Spanish and ensure the equivalence of items (Gibson, 1999). Three certified translators and a group of specialists from bilingual teams in the United States and Spain were involved in this process. The materials forming the test were first translated into Spanish by the three translators. The Spanish version of the test was then translated back into English by the bilingual specialists. Finally, the back-translation was reviewed by the research team to ensure that the meaning of the items was consistent in the different translations. As a final result, we obtained a scale in Spanish equivalent to the English-language original.

**Participants and procedure.** The participants in the study were 135 students from different disciplines (78% information technology engineering and 22% psychology) at a public university in Madrid. Men made up 71% of the total sample, and the average age was 23 years ( $SD = 0.93$ ). All of the participants were from Spain and had Spanish as their mother tongue.

Participation in the study was mandatory for the students, being part of a program of practices in different subjects. Participants signed a consent form regarding their participation in the study. They could voluntarily decline to be enrolled in the research by not providing their records without any adverse effect on their class grades. At the beginning of the academic semester, the researchers administered the TWKSAT to the students as a part of their class requirements. The objectives and reasoning behind the study were explained by the professors at the end of the course. Each participant was also given a brief report with the scores ob-

<sup>1</sup> The TWKSAT is a commercial test and we therefore do not provide the content of the items included in the test so as to protect the intellectual property rights of its authors. The test can be acquired at <http://www.vangent-hcm.com/Solutions/SelectionAssessments/SkillsAbilitiesAssessments/>. We use the original numbering of the items in our presentation of results to facilitate understanding.

tained on the test and the meaning of each of the teamwork competencies.

The validity of the TWKSAT content was analyzed by three experts in the field of teamwork. All raters had more than 10 years' experience in conducting team-based research and consulting. The experts analyzed the TWKSAT items as described in the Results section.

## Results

**Description of items.** Table 1 shows the mean (difficulty index), standard deviation,

asymmetry, kurtosis, and discrimination index for each item on the scale, as well as the Cronbach's alpha estimated for all 35 items together. As shown, the range of difficulty varies between .08 (Item 26) and .93 (Item 6). With regard to the discrimination index, four items (7, 12, 27, and 35) are negatively correlated with the scale and another series of items (2, 4, 10, 11, 15, 17, 24, and 25) display a positive but very low adjusted correlation (<.15). The total alpha for the scale is .60. However, if the negatively correlated items and those with a correlation of

Table 1  
*Descriptive Statistics for Study 1 Teamwork Knowledge, Skills, Attitudes Test Original Items*

Item	Mean	SD	Discrimination index (corrected item-total correlation)	Skewness	Kurtosis
01	0.59	0.49	.22	-0.80	-1.34
02	0.39	0.49	.03	0.68	-1.53
03	0.79	0.41	.18	-1.62	0.63
04	0.68	0.47	.01	-1.44	0.08
05	0.53	0.50	.17	-0.23	-1.94
06	0.93	0.25	.35	-5.39	27.02
07	0.49	0.50	-.07	0.02	-1.99
08	0.91	0.29	.37	-4.08	14.60
09	0.34	0.47	.19	0.25	-1.93
10	0.72	0.45	.13	-1.26	-0.40
11	0.50	0.50	.14	-0.46	-1.78
12	0.29	0.45	-.01	0.70	-1.49
13	0.71	0.46	.15	-1.01	-0.96
14	0.29	0.45	.22	0.69	-1.51
15	0.70	0.46	.10	-0.88	-1.22
16	0.11	0.32	.30	0.60	-1.63
17	0.69	0.46	.02	-0.79	-1.37
18	0.16	0.37	.31	0.15	-1.97
19	0.70	0.46	.30	-0.86	-1.25
20	0.76	0.43	.35	-1.62	0.63
21	0.56	0.50	.21	-1.06	-0.85
22	0.61	0.49	.29	-0.60	-1.63
23	0.50	0.50	.27	-0.33	-1.88
24	0.40	0.49	.01	-0.02	-1.99
25	0.32	0.47	.02	0.43	-1.80
26	0.08	0.28	.24	0.92	-1.14
27	0.38	0.49	-.06	0.01	-1.99
28	0.68	0.47	.36	-1.18	-0.58
29	0.22	0.41	.15	1.15	-0.67
30	0.15	0.36	.41	0.13	-1.98
31	0.64	0.48	.15	-1.00	-0.99
32	0.09	0.29	.34	0.31	-1.89
33	0.61	0.49	.12	-0.40	-1.83
34	0.23	0.42	.10	0.76	-1.41
35	0.50	0.50	-.05	0.02	-1.99

Note.  $N = 135$ .  $\alpha = .60$ .

less than .15 are discarded, the alpha scale rises to .71.

**Content analysis.** To examine the validity of the test content, we employed the strategy proposed by Rovinelli and Hambleton (1977) in which a group of experts expresses the degree to which each item measures each of the previously defined evaluation objectives. Accordingly, three experts separately analyzed the 35 items composing the test, indicating the extent to which each item was representative of each of the teamwork subcompetencies defined by Stevens and Campion (1994) and relevant to the measurement of the target. A score of 1 was assigned if the item was considered appropriate, -1 if not, and 0 if the expert in question was unsure. The resulting index will be 1 when all of the experts concur that an item is appropriate in view of the target measured. The experts were also asked to make any general comments on the measurement items included.

Table 2 summarizes the results. A number of items display a high level of item-target congruence ( $>.80$ ), other items show a congruence index of less than .50 for all of the measurement targets (27, 29, and 33), and Item 31 falls short of the threshold value of .30 for inclusion in the table (Thorn & Deitz, 1989). Table 2 also provides valuable information related to the degree to which the test items meet the proposed measurement objectives. There are no items associated with the subcompetencies A2 (“Recognize the type and source of conflict”), C1 (“Understand and use communication networks”), and D2 (“Monitor, assess and provide feedback on individual and group performance”). Moreover, the number of items that are congruent with the different measurement targets varies between a single highly congruent item (23) for A3 (“Use win-win strategies”) and four items (1, 4, 11, and 19) for D1 (“Help set specific, challenging and accepted objectives”).

Finally, all of the experts remarked in their comments that the test items allow assessment of only the respondents’ knowledge, although the original Stevens and Campion (1994) model refers to “teamwork knowledge, skills and abilities.”

**Dimensionality analysis.** A confirmatory factor analysis ( $N = 135$ ) was carried out on the correlation matrix to examine the dimensionality of the test. As these are dichotomous (right/wrong) items, the tetrachoric correlation matrix

was analyzed using Mplus (Muthén & Muthén, 2006). Following the substantive model proposed by Stevens and Campion (1994, 1999), we tested five models: (a) five-factor orthogonal model, (b) two-factor orthogonal model, (c) five-factor correlated model, (d) two-factor correlated model, and (e) one-factor model.

Table 3 presents the fit indices established for the different estimated models. The orthogonal solutions of the five-factor and two-factor models do not fit the data. Values in each index are substantially below the usual standards. However, when correlation between factors is allowed, the model fits the data better. Even so, the only adjustment index for which a good fit is obtained is the root mean square error of approximation ( $<.05$ ). A similar effect is observed with a single-factor model.

In addition, to deeply explore the dimensionality of the items proposed, we conducted an exhaustive exploratory factor analysis of the different factorial solutions from one to seven factors, including a review of both orthogonal and oblique rotations. The results do not allow us to clearly label the factors according to the Stevens and Campion model (1994).

**Criterion validity analysis.** To provide further evidence of the criterion-related validity of the TWKSAT, we correlated the scores in the TWKSAT against the criterion for individual performance in a team task. A subsample of the study ( $n = 30$ ; 56% women; average age = 23 years) formed by the participants in an undergraduate teamwork course was observed during the resolution of a group decision-making task. The participants were randomly assigned to five-member teams. The task required the teams to generate effective measures to resolve traffic problems caused by improper parking on a university campus.<sup>2</sup> The teams were allowed 30 min to complete the task and were video-recorded for subsequent analysis. Both team tasks and video recording were standard features of the course, and the participants were therefore familiar with these procedures. An ad hoc code (see Appendix A) was designed for behavioral observation. The code categories describe specific behaviors associated with the five teamwork competencies identified by Stevens and Campion (1994, 1999; e.g., “Request

<sup>2</sup> The task materials are available on request.

Table 2  
*Teamwork Knowledge, Skills, Attitudes Test Congruence Indexes*

Item	A1	A2	A3	B1	B2	C1	C2	C3	C4	C5	D1	D2	E1	E2
06	<b>0.88</b>						0.35							
07	<b>0.65</b>						0.47							
23			<b>0.92</b>											
10				<b>0.55</b>				0.37						
12				<b>0.64</b>										
26				<b>0.65</b>										
22					<b>0.97</b>									
27					<b>0.31</b>									
28					<b>0.92</b>									
29					<b>0.31</b>									
15							<b>0.96</b>							
32							<b>0.79</b>							
33							<b>0.45</b>							
34							<b>0.64</b>							
09								<b>0.58</b>						
16								<b>0.94</b>	0.76					
30								<b>0.60</b>						
14							0.58		<b>0.94</b>					
24								0.40	<b>0.94</b>					
35							0.59		<b>0.95</b>					
02										<b>0.79</b>				
17										<b>0.99</b>				
01											<b>0.95</b>			
04											<b>0.92</b>			
11											<b>0.97</b>			
19											<b>1.00</b>			
18											<b>0.62</b>			
08													<b>0.88</b>	0.71
13													<b>0.62</b>	
20													<b>0.78</b>	
21													<b>0.95</b>	0.41
25													<b>0.55</b>	0.37
05													<b>0.59</b>	
03													0.71	<b>0.88</b>
31														

*Note.* A1 = The knowledge, skills, and attitudes (KSA) to recognize and encourage desirable, but discourage undesirable, team conflict; A2 = The KSA to recognize the type and source of conflict confronting the team and to implement an appropriate conflict resolution strategy; A3 = The KSA to employ an integrative negotiation strategy rather than the traditional distributive strategy; B1 = The KSA to identify situations requiring participative group problem solving and to utilize the proper degree and type of participation; B2 = The KSA to recognize the obstacles to collaborative group problem solving and implement appropriate corrective actions; C1 = The KSA to understand communication networks and to utilize decentralized networks to enhance the communication where possible; C2 = The KSA to communicate openly and supportively; C3 = The KSA to listen nonevaluatively and appropriately use active listening techniques; C4 = The KSA to maximize consonance between nonverbal and verbal messages, and to recognize and interpret the nonverbal messages of others; C5 = The KSA to engage in ritual greetings and small talk, and a recognition of their importance; D1 = The KSA to help establish specific, challenging, and accepted team goals; D2 = The KSA to monitor, evaluate, and provide feedback on both overall team performance and individual team member performance; E1 = The KSA to coordinate and synchronize activities, information, and task interdependencies between members; E2 = The KSA to help establish task and role expectations of individual team members and to ensure proper balancing of workload in the team. To facilitate reading the indexes, indexes  $< .30$  have been eliminated; indexes  $> .80$  appear in bold italic type; and the greater index obtained for each item appears in bold type.

additional information from team members”). Two groups of five judges independently analyzed the recordings, noting the frequency of teamwork behaviors associated with a particular

competency displayed by each of the team members. All of the judges were postgraduate students who were blind to the objectives of the study and independent of the research group.



Table 3  
*Model Fit Indexes*

Model	RMSEA	CFI	TLI
Five factor (orthogonal)	.082	.067	.008
Two factor (orthogonal)	.07	.324	.282
Five factor (correlated)	.04	.776	.762
Two factor (correlated)	.038	.807	.792
One factor	.041	.774	.76

$N = 135$ ; RMSEA = root mean square error of approximation; CFI = confirmatory fit index; TLI = Tucker–Lewis fit index.

They were trained in the use of the code and underwent a trial period to align criteria and discuss inconsistencies. The mean interjudge reliability for the different competencies was acceptable ( $\kappa = .87$ ); therefore, we computed the mean of their assessments to obtain a single score for each competency. Finally, we computed the mean of the scores for each individual and the five competencies, as in the case of the questionnaire, to obtain a total teamwork competency score. As expected, the TWKSAT score was positively correlated with the observed measure of teamwork skills ( $r = .43, p < .05$ ).

## Discussion

The first study revealed certain weaknesses in the TWKSAT (Stevens & Campion, 1999) that affect its use in academic and professional contexts. Specifically, the following deficiencies were observed in three different facets: (a) Some items show very low reliability indices affecting the general reliability of the scale, (b) the contents of the original model proposed by Stevens and Campion (1994) are not fully represented in the test, and (c) the dimensional structure obtained from the factor analysis is not well aligned with the substantive model.

With regard to the first weakness, the results of the study show that certain items should be eliminated from the scale, as their adjusted correlations with the total for the scale are less than the generally accepted standards.

Turning to the second weakness, the content analysis indicates that certain teamwork subcompetencies identified in the general model are not picked up by the test items. Thus, no items exist associated with the subcompetencies A2 (“Recognize the type and source of conflict”), C1 (“Understand and use communica-

tion networks”), and D2 (“Monitor, assess and provide feedback on individual and group performance”). The content analysis also revealed that certain items are not associated with any of the proposed measurement targets (Items 27, 29, 31, and 33). In addition, the format of the items is designed to measure “knowledge,” but not skills or aptitudes.

Finally, Stevens and Campion (1994) established 14 teamwork subcompetencies in their model, grouped into five competencies, which are, in turn, integrated in two dimensions. However, our results indicate a structure that tends toward unidimensionality.

To sum up, the results of this study provide an explanation of the low reliability indices reported in previous research and suggest how a more reliable measurement could be obtained from the TWKSAT without affecting its predictive capacity. In line with prior studies using the original English version of the TWKSAT (Chen et al., 2004; McClough & Rogelberg, 2003; Stevens & Campion, 1999), the Spanish adaptation of the test used in our study reveals similar correlation indices ( $r = .43, p < .05$ ).

It is necessary to recognize that our results may be influenced by cultural differences between U.S.-based samples used by Stevens and Campion (1999) to develop and validate the TWKSAT and our Spain-based sample. Cross-cultural studies (e.g., Earley & Erez, 1997; Hofstede, 1980, 1983a, 1983b) show that cultural differences may affect the processes and outcomes of individuals and groups in organizations (Cox, Lobel, & McLeod, 1991; Gibson, 1999). Thus, in contrast to North Americans, Spaniards tend to be (a) higher in collectivism, so they tend to put group interests first (Triandis, 1995) and look more actively for social acceptance, strong group identity, and the development of personal relationships (Grimm, Church, Katigbak, & Reyes, 1999); (b) higher in power distance, leading them to behave more submissively with managers and avoid disagreements and feeling more comfortable working in teams with strong directivity (Earley & Erez, 1997); (c) lower in action orientation (Maznevski, DiStefano, Gomez, Noorderhaven, & Wu, 1997), which could make them work less hard (Hampden-Turner & Trompenaars, 1993) and be more resistant to work by objectives (Kluckhohn & Strodtbeck, 1961); and (d) more likely to believe that external forces determine

their successes and failures (Trompenaars, 1993). According to these cultural differences between Spanish and North American populations, the way employees approach their professional performance in teams and, therefore, the way employees respond to the TWKSAT may be different. Thus, generalization of our findings to the context of teamwork in North American cultures should be done with caution.

To conclude, our findings suggest the need to improve the measurement of teamwork competencies. The TWKSAT allows measurement of only the General Teamwork competency, which constricts examination of the conceptual richness expressed in Stevens and Campion's (1994) original model. Consequently, it is desirable to develop new items capable of capturing the different subcompetencies, as well as obtaining more reliable scores in the five competencies.

## Study 2

The second study analyzed a new measure of the teamwork knowledge, skills, and abilities based on Stevens and Campion's (1994) model, which is called the Teamwork Competency Test (TWCT). For this purpose, we developed and adapted new items in light of the findings from Study 1. The new test comprises 36 items, which were administered to a sample of 120 university students to analyze the functioning of the items and to test the TWCT's reliability, content validity, and dimensionality.

## Method

**Item development for TWCT.** Based on results from Study 1, we developed new items to represent the 14 subcompetencies established by Stevens and Campion (1994). The items were worded in the observable behavior format. An initial set of 83 items was constructed, which was progressively refined to the final 36 items. Based on interviews with three experts in the fields of teamwork and organizational behavior, several items were rephrased (seven items) or left out (31 items). The remaining 52 items were applied to different samples and, taking into account the item statistics (corrected item-total correlation and factor loadings), 16 items were deleted. Using the selected 36 items, we conducted a pilot study on 26 members of

four software development programmers' teams from a small information technology company. All team members responded to the questionnaire and provided feedback on the items. Most items were understood and perceived as unambiguous. Some minor changes were made based on their comments, but all the items were retained. As a result, the new measure contains 36 items drafted in Spanish (see the complete scale in Appendix B) and uses a 4-point response scale of frequency (1 = *never/almost never* and 4 = *always/almost always*).

**Participants and procedure.** The sample comprised 120 final-year psychology students at a large Spanish public university, 68.7% of whom were women. The average age was 23 years ( $SD = 0.96$ ). The procedure was identical to that in Study 1. The assistance of the same experts was used to analyze the validity of the test content as in Study 1.

## Results

**Description of items and scales.** Table 4 summarizes the main descriptive statistics for the questionnaire items. The mean scores for all items were above 2.5 (the theoretical midpoint on the response scale), except those related to the Collaborative Problem Solving dimension, in which the averages were somewhat lower (between 1.94 and 2.28). The reliability indices for both the scales and the full questionnaire (.89) were satisfactory.

**Content analysis.** Table 4 presents the congruence indices (Rovinelli & Hambleton, 1977) for each item based on the subcompetencies proposed by Stevens and Campion (1994). The congruence indices for all of the items were satisfactory (minimum = 0.56, maximum = 1.00). Finally, the whole content domain established in the model was represented by the items developed.

**Dimensionality analysis.** We carried out various factor analyses to explore the dimensionality of the questionnaire above and beyond the adequacy and relevance of the items. Given the high means for the items and the deviation of distributions away from normal, we used the minimum unweighted least squares method for factor extraction, employing the FACTOR program (Lorenzo-Seva & Ferrando, 2006). Bartlett's sphericity test ( $\chi^2 = 5557.6$ ,  $df = 703$ ,  $p < .001$ ) and the Kaiser-Meyer-Olkin index

Table 4  
*Teamwork Competency Test Item Statistics, Congruence Indexes, and Scale Reliabilities*

Item	<i>M</i>	<i>SD</i>	Content														
			A1	A2	A3	B1	B2	C1	C2	C3	C4	C5	D1	D2	E1	E2	
1	3.34	0.67	1.00														
16	3.00	0.75	1.00														
19	2.91	0.74	0.56							0.83							
21	3.12	0.75	0.77							0.63						0.5	
7	3.41	0.60		0.69							0.83						
10	2.47	0.89		0.64													
3	2.88	0.79			0.94												
15	3.27	0.68			0.81					0.81							
11	3.00	0.86				0.81											
25	3.04	0.84				1.00											
14	2.71	0.74					0.70										
24	2.51	0.81					0.83										
26	2.61	0.76					1.00										
5	2.83	0.70						0.95									
17	3.27	0.70							0.95								
28	3.28	0.69							1.00								
29	3.31	0.68							0.69								
2	3.43	0.61								0.78							
27	3.37	0.63								1.00							
8	3.57	0.61									0.89						
9	3.13	0.75									0.83						
12	3.25	0.79										1.00					
30	3.23	0.78										1.00					
35	2.70	0.75											1.00				
6	2.99	0.76												1.00			
20	3.03	0.81													0.87		
22	2.91	0.83													0.95		
31	3.00	0.77													1.00		
32	3.06	0.78													1.00		
36	2.86	0.80													0.95		
4	3.06	0.72														0.87	
23	3.33	0.75														1.00	
33	3.11	0.68														0.87	
34	3.07	0.68														0.87	
13	3.65	0.60															0.89
18	3.05	0.73															1.00
Total	112.9	12.00															

*Note.* A1 = The knowledge, skills, and attitudes (KSA) to recognize and encourage desirable, but discourage undesirable, team conflict; A2 = The KSA to recognize the type and source of conflict confronting the team and to implement an appropriate conflict resolution strategy; A3 = The KSA to employ an integrative negotiation strategy rather than the traditional distributive strategy; B1 = The KSA to identify situations requiring participative group problem solving and to utilize the proper degree and type of participation; B2 = The KSA to recognize the obstacles to collaborative group problem solving and implement appropriate corrective actions; C1 = The KSA to understand communication networks and to utilize decentralized networks to enhance the communication where possible; C2 = The KSA to communicate openly and supportively; C3 = The KSA to listen nonevaluatively and appropriately use active listening techniques; C4 = The KSA to maximize consonance between nonverbal and verbal messages, and to recognize and interpret the nonverbal messages of others; C5 = The KSA to engage in ritual greetings and small talk, and a recognition of their importance; D1 = The KSA to help establish specific, challenging, and accepted team goals; D2 = The KSA to monitor, evaluate, and provide feedback on both overall team performance and individual team member performance; E1 = The KSA to coordinate and synchronize activities, information, and task interdependencies between members; E2 = The KSA to help establish task and role expectations of individual team members and to ensure proper balancing of workload in the team.

(.66) confirmed that the item correlation matrix could be factorized. The factorial solution was obliquely rotated. Eleven factors were found to have an eigenvalue of more than 1.00 after extraction. The parallel analysis retained only the first eight factors extracted. In line with this analysis, the eight-factor model was the one that best reflected the initial substantive model, explaining 56% of the total variance, with only 18% residual errors of more than .05. The eight-factor solution was compared with a five-factor solution, as proposed in the substantive model. The percentage variance explained was found to be lower (43%), as was the percentage of residual errors greater than .05 (33%). In addition, the residual mean squares error average showed a marginally acceptable value (.085) in the eight-factor model, but a clearly unacceptable value in the five-factor model (.11). Therefore, it appears that the eight-factor model better reproduces the analyzed data matrix compared with the five-factor model.

Table 5 presents the results obtained in the configuration matrix after oblique rotation. Although this matrix does not exactly reproduce the expected structure of 14 subcompetencies, a detailed analysis indicates that it reflects the substantive model—with some variations—on which the development of the items was based. The first factor extracted represents the Conflict Resolution competency. The items developed in relation to this competency (especially those referring to the recognition of conflict: Items 1, 10, 15, 16, 19, and 21) are associated with this factor, as are the Communication competency items related to open communication and support. They are joined by Item 18 (designed for the Planning and Coordination competency). The second factor extracted reflects the Planning and Coordination competency (Items 4, 13, 23, 33, and 34). This factor also includes Item 8 (designed for the Communication competency). The items for the Goal Setting and Performance Management competency appear in Factors 3 and 6. Items related to “monitoring, assessing and providing feedback on individual and group performance” (Items 6, 22, 31, and 35) are associated with Factor 3, and those referring to “offering teammates feedback on their results” (Items 20, 32, and 36) are associated with Factor 6. Factors 4 and 5 reflect the Collaborative Problem Solving competency. Factor 4 is associated with items referring to

“recognizing obstacles to participative problem solving” (Items 14, 24, and 26), and Factor 5 is associated with items referring to “identifying situations that require participation in decision making” (Items 11 and 25). Finally, Factors 7 and 8 reflect the Communication competency. The items related to “active listening, nonverbal communication and recognition of communication networks” (Items 2, 3, 5, 7, 9, and 27) are associated with Factor 7, and those related to “informal communication acts with other team members” (Items 12 and 30) are associated with Factor 8.

Finally, we performed a confirmatory factor analysis ( $N = 120$ ) on the scores obtained for the eight factors to verify whether the eight factors extracted are associated with two general dimensions, as proposed in the substantive model, or with a single dimension as Study 1 suggests. The following factors were assigned to the Interpersonal competencies dimension in the two-dimensional model (see Figure 2): 1 (Conflict Resolution), 4 (Collaborative Problem Solving: Group), 5 (Collaborative Problem Solving: Individual), 7 (Communication: Active Listening), and 8 (Communication: Informal). The following factors were assigned to the Self-Management dimension: 2 (Planning), 3 (Performance Objective Management: Monitoring), and 7 (Performance Objective Management: Feedback).

In view of the weightings assigned to the different variables, all of the parameters estimated in the two-factor model are significant ( $p < .001$ ), except for the relationship between Factor 4 and Factor 5 and the Interpersonal Relations dimension ( $p = .001$  and  $p = .040$ , respectively). All of the parameters estimated are again significant ( $p < .001$ ) for a one-factor model, except for the relationship between Factor 4 and Factor 5 and the General Teamwork factor ( $p = .002$  and  $p = .042$ , respectively). None of the standardized residuals attains a value of  $|\text{res}_z| > 2$  for the one-factor model. However, a high value,  $|\text{res}_z| > 2$ , appears in the two-factor model for the relationship between Factor 1 and Factor 7, although this is not an especially large deviation.

The values obtained for the goodness-of-fit estimators for both models indicate a good fit in both models. For the one-factor model and two-factor model, the values were, respectively,  $\chi^2 = 0.181$  and  $0.242$ , standardized root mean

Table 5  
Exploratory Factor Analysis for Teamwork Competency Test Items Configuration Matrix

Item	Factor							
	1	2	3	4	5	6	7	8
28	<b>0.93</b>	-0.09	-0.05	-0.02	-0.10	-0.02	-0.03	0.05
17	<b>0.92</b>	-0.12	-0.02	-0.05	-0.07	-0.02	-0.04	0.08
29	<b>0.90</b>	-0.11	-0.03	-0.05	-0.10	-0.06	-0.06	0.12
15	<b>0.47</b>	0.08	0.05	0.04	0.16	0.07	0.02	-0.04
19	<b>0.35</b>	0.03	0.05	0.06	0.09	0.16	0.19	-0.03
18	<b>0.29</b>	0.00	0.25	0.09	0.03	0.06	0.17	0.03
01	<b>0.24</b>	0.07	0.12	0.10	0.24	0.14	0.16	-0.02
10	<b>0.23</b>	-0.02	0.15	0.08	0.09	0.06	0.14	-0.05
21	<b>0.21</b>	-0.05	0.06	0.13	0.04	<b>0.29</b>	-0.03	-0.08
16	<b>0.18</b>	0.13	<b>0.20</b>	-0.01	0.13	0.15	0.09	0.05
34	0.05	<b>-0.94</b>	-0.02	0.01	-0.03	0.10	-0.06	-0.03
33	0.02	<b>-0.91</b>	-0.04	-0.03	-0.04	0.06	-0.03	-0.01
04	0.05	<b>-0.84</b>	0.03	0.02	0.01	0.09	-0.07	-0.05
23	0.01	<b>-0.55</b>	0.13	-0.05	0.05	0.06	0.12	-0.02
08	0.04	<b>-0.22</b>	0.01	0.01	0.05	-0.07	0.11	0.08
13	0.00	<b>-0.19</b>	0.13	-0.02	0.05	-0.02	0.18	0.06
31	-0.06	-0.05	<b>0.93</b>	-0.03	-0.02	0.01	-0.06	0.03
06	-0.06	0.01	<b>0.91</b>	0.00	-0.02	0.02	-0.05	0.01
22	0.00	-0.05	<b>0.87</b>	0.02	-0.01	-0.01	-0.03	-0.07
35	<b>0.12</b>	-0.01	<b>0.11</b>	0.10	0.03	0.05	-0.01	0.08
26	-0.02	-0.01	-0.01	<b>0.95</b>	-0.06	0.03	-0.07	0.00
14	-0.02	0.01	-0.09	<b>0.89</b>	-0.07	-0.02	-0.06	0.01
24	-0.12	0.03	0.08	<b>0.80</b>	0.05	-0.02	0.03	0.01
11	-0.08	-0.02	-0.03	-0.02	<b>0.98</b>	-0.03	-0.08	0.02
25	-0.07	-0.02	-0.03	-0.05	<b>0.93</b>	-0.03	-0.10	0.03
20	-0.06	-0.10	-0.02	-0.01	-0.04	<b>0.89</b>	0.03	0.02
32	-0.09	-0.14	-0.02	-0.05	-0.01	<b>0.85</b>	0.03	0.08
36	0.11	0.06	0.13	0.04	0.00	<b>0.53</b>	0.08	0.09
27	-0.08	-0.06	-0.01	-0.04	-0.07	0.09	<b>0.90</b>	-0.08
02	-0.10	-0.05	-0.02	-0.06	-0.10	0.05	<b>0.79</b>	-0.03
03	0.16	0.04	-0.08	0.00	0.03	0.13	<b>0.39</b>	0.03
07	0.03	0.03	0.03	0.09	0.10	0.02	<b>0.36</b>	0.15
05	0.09	-0.03	0.14	0.01	-0.07	-0.08	<b>0.25</b>	0.09
09	0.08	-0.16	-0.03	0.17	0.11	-0.11	<b>0.25</b>	0.00
30	-0.02	0.01	-0.05	0.01	0.03	0.03	-0.03	<b>1.00</b>
12	0.02	0.04	-0.01	0.01	0.00	0.07	-0.06	<b>0.90</b>

Note. *N* = 120; Extraction = unweighted least squares; Rotation = Oblimin–Kaiser; Convergence = 12 iterations. Items rationally linked with the factor are in bold type. Items not rationally linked with the factor and up to .10 value are in bold italic type.

residual = .048 and .047, root mean square error of approximation = .037 and .031, comparative fit index = .976 and .983, and Tucker–Lewis index = .966 and .975 (Figure 2).

**Discussion**

Overall, the results of the second study show that the TWCT reasonably covers the whole content domain proposed by Stevens and Cam- pion (1994), is reliable, and its dimensional structure adequately reflects the original sub-

stantive model (although it does not do so exactly). With regard to the first point, the inter- judge analysis indicates that the TWCT items are representative of the 14 subcompetencies defined by the authors. In terms of reliability, the five scales present adequate Cronbach’s al-phas of more than .80, except the Conflict Res-olution scale, which has an alpha of .71. The alpha for the total scale is also adequate (.89). Finally, the dimensionality analysis reveals a latent structure for the TWCT that is substan-

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

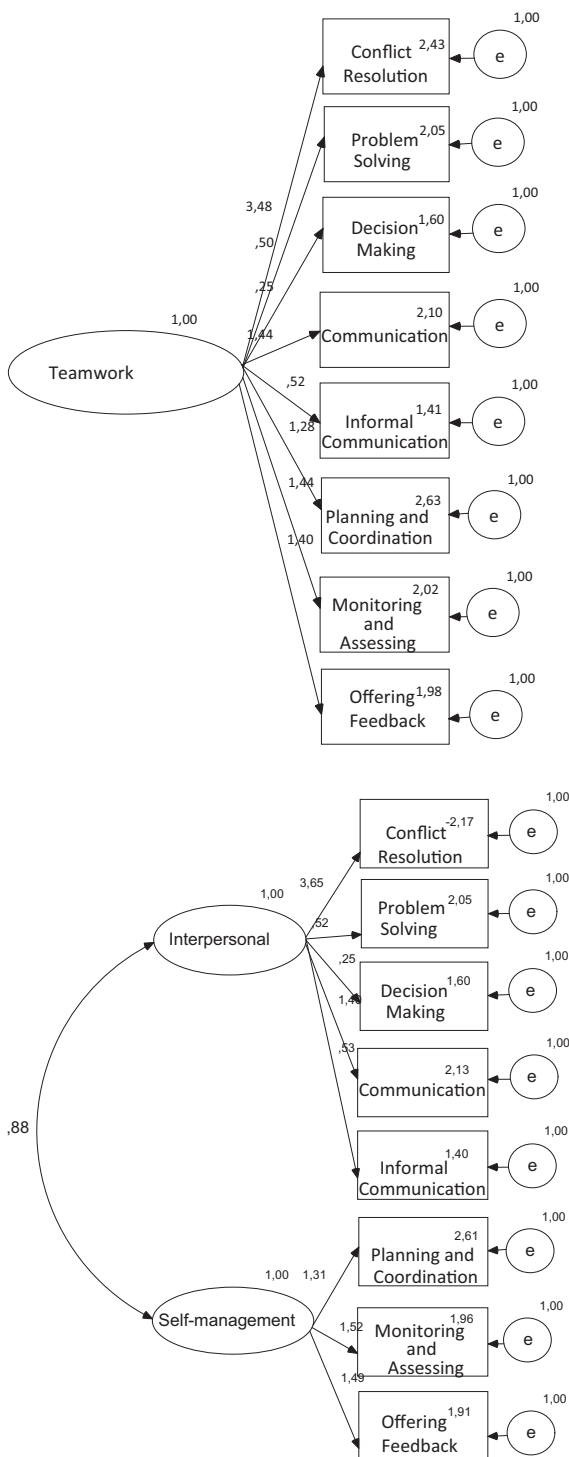


Figure 2. Confirmatory factor analysis solutions for the one-factor and two-factor models.

tially associated with Stevens and Campion's (1994) model. However, the structure observed does not exactly reproduce either the structure of the 14 subcompetencies or the structure of the five competencies.

Our results suggest the need to reconfigure the originally designed scales. This is particularly important regarding the Conflict Resolution competency. The analyses of both dimensionality and congruence reveal a strong association between the Conflict Resolution items and the Communication items (related to open communication and support). It seems reasonable to integrate the items in the Communication subcompetency into the Conflict Resolution competency, as the factor analysis indicates. This is also true with regard to Item 18, which both the judges and the factor analysis associate with that competency. The Conflict Resolution scale with an alpha of .84 would thus be established with the same items as originally designed. The Collaborative Problem Solving scale is also maintained intact, integrating the Factor 4 and Factor 5 items with an alpha of .83. The Communication scale contains the Factor 7 and Factor 8 items associated with communication items, except for open communication and support, obtaining an alpha of .72. The Objective Management and Performance competency would be formed by all of the items from Factors 3 and 6 with an alpha of .82. Finally, the Planning and Coordination competency integrates all of the items associated with Factor 2 ( $\alpha = .88$ ), except Item 18, which is moved to Conflict Resolution.

### Study 3

The third study had a dual objective: (1) to obtain initial evidence for the convergent validity of the new measure of teamwork competency developed in Study 2, and (2) to compare the new measure with the original test proposed by Stevens and Campion (1999). A sample of employees was used to analyze the relationship between the TWCT and (a) the original TWKSAT, (b) supervisor assessments of employees' teamwork competencies, and (c) self-assessments of competency by employees themselves.

### Method

**Sample.** A total of 91 employees working in a major Spanish power utility took part in the study. Men made up 81% of the total sample, and the average age was 29.6 years ( $SD = 3.38$ ). The measurements were made as part of the activities undertaken by the employees within the framework of a mandatory skills development program set up by the company. Participants signed a consent form regarding their participation in the study. They could decline to be enrolled in the research by not providing their records without any adverse effect on their participation in the skills development program. We also obtained company consent to use the data in the research.

#### Measures.

**TWKSAT.** We administered the original version of the Stevens and Campion (1999) test used in Study 1.

**TWCT.** The version of the test designed in Study 2 was administered. Six different measures were established for each participant, consisting of a total score and a score for each of the five competencies.

#### Team performance: Supervisor assessment.

The assessments of each employee's immediate superior were obtained specifically for the purposes of the study and did not form part of the company's usual performance management process. The information gathered on the employees was strictly training-related. The supervisor questionnaire contained eight items describing different teamwork behaviors, to which they were asked to respond on a 4-point scale (1 = *completely disagree*, 4 = *completely agree*). An example item is "He or she (the employee) participates actively in work meetings (giving his or her opinion, asking questions, etc.)." The items were drawn from a competency model used in the company.

#### Team performance: Self-assessment.

The same eight-item questionnaire expressed in the first person was also answered by each employee to assess his/her usual behavior in work team situations.

**Procedure.** The supervisors' assessments (Measure 3) and the self-assessments of the employees (Measure 4) were obtained concurrently at the beginning of the course. The supervisors gave their responses online via a

user/key connection to a web page containing the questionnaire and instructions for completion. The TWKSAT and TWCT measures (respectively, Measures 1 and 2) were collected at a first training session 1 week later. Both questionnaires were administered in a paper-and-pencil format, counterbalanced to avoid effects associated with the order of presentation.

## Results

Table 6 summarizes the descriptive statistics and intercorrelations for all of the variables in the study. The reliability of the two criteria measures used was .70 for team performance supervisor evaluation and .54 for team performance self-evaluation. The two versions of the TWKSAT were positively correlated ( $r = .47, p < .01$ ), as were the assessments made by the supervisors and the employees ( $r = .25, p < .05$ ). As in Study 2, the competencies evaluated in the TWCT were positively correlated (except Competency B: Collaborative Problem Solving). Also, the competencies as captured by the TWCT were positively correlated with the original TWKSAT, except Competency B.

The supervisor assessment correlation with the TWCT was .34 ( $p < .01$ ) compared with a correlation of .26 ( $p < .05$ ) with the TWKSAT. These correlations are not statistically different ( $T = -0.76, p < .05$ ). A similar effect was observed in the employee self-assessments, which showed a correlation of .39 ( $p < .01$ ) with the TWCT and .21 ( $p < .05$ ) with the TWKSAT. These correlations are statistically

different ( $T = -1.78, p < .05$ ). Analysis of the TWCT competencies shows that Competencies A (Conflict Resolution), D (Goal Setting and Performance Management), and E (Planning and Coordination) are positively correlated both with the supervisor's assessment ( $r = .24, p < .05$ ;  $r = .36, p < .01$ ;  $r = .27, p < .01$ , respectively) and with employees' self-assessments ( $r = .30, p < .01$ ;  $r = .38, p < .01$ ;  $r = .37, p < .01$ , respectively).

Finally, we performed a hierarchical regression analysis examining the proportion of the variance in the scores obtained from the supervisors' assessments due to the TWKSAT and TWCT, respectively. In the first step, the TWKSAT score was introduced, and the TWCT score was introduced in a second step. Table 7 indicates that the percentage of explained variance increases significantly at 6.3%,  $\Delta R^2 = .063, p < .05$ ;  $F(2.88) = 6.60$ ;  $\beta = .285, p < .05$ , when the TWCT score is introduced.

## Discussion

The results of the third study provide favorable empirical evidence for the convergent validity of the TWCT. The total score in this new version of the test is positively correlated with the other measures in the study, namely (a) the score in the original Stevens and Campion test (1999), (b) the supervisors' assessments of the teamwork behavior of employees in their routine work, and (c) the self-assessments made by the employees themselves. Furthermore, the results indicate that the Conflict Resolution, Goal Setting and Performance Management, and

Table 6  
Descriptive Statistics and Intercorrelations for Study 3 Measures

Measure	<i>M</i>	<i>SD</i>	$\alpha$	1	2	3	4	5	6	7	8
1. TWKSAT	21.9	2.3	.37								
2. TWCT Total	109.0	10.9	.84	.47**							
3. TWCT KSA_A	30.5	4.6	.84	.36**	.81**						
4. TWCT KSA_B	13.6	2.6	.26	.10	.42**	.07					
5. TWCT KSA_C	25.0	2.8	.35	.37**	.59**	.39**	.00				
6. TWCT KSA_D	19.9	3.3	.66	.41**	.81**	.59**	.24*	.33**			
7. TWCT KSA_E	19.9	2.4	.55	.33**	.72**	.43**	.36**	.29**	.51**		
8. Team performance supervisor evaluation	16.2	3.3	.70	.26*	.34**	.24*	.13	.16	.36**	.27**	
9. Team performance self-evaluation	23.4	2.8	.54	.21*	.39**	.30**	.08	.19	.38**	.37**	.25*

Note. TWKSAT = Teamwork Knowledge, Skills, Attitudes Test; TWCT = Teamwork Competency Test;  $N = 91$ ; KSA\_A = Conflict Resolution; KSA\_B = Collaborative Problem Solving; KSA\_C = Communication; KSA\_D = Goal Setting and Performance Management; KSA\_E = Planning and Task Coordination.

\*  $p < .05$ . \*\*  $p < .01$ .



Table 7  
*Hierarchical Regression Analysis Over the Team Performance Supervisor Evaluation*

Step	<i>df</i>	<i>F</i>	$\Delta R^2$	$\beta$
1	1, 89	6.42	.067*	
TWKSAT				.259*
2	2, 88	6.60	.063*	
TWKSAT				.125
TWCT				.285*

Note. TWKSAT = Teamwork Knowledge, Skills, Attitudes Test; TWCT = Teamwork Competency Test.

\*  $p < .05$ .

Planning and Coordination competencies are the only competencies of the five contained in the TWCT that correlate independently with the criterion variables.

Contrary to our expectations, however, the correlations between the different subscales in the test are not all either significant or positive. Specifically, Collaborative Problem Solving is weakly related to the other dimensions except Goal Setting and Performance Management and Planning and Coordination.

Another interesting feature is the significant, positive correlation found between the original TWKSAT and the supervisors' assessments. This is in line with the original study carried out by Stevens and Campion (1999). Finally, comparison of the predictive capacity of the two versions of the TWKSAT suggests that the TWCT is a better predictor of the self-assessment of teamwork—but not of the supervisors' assessments—than the original test.

### General Discussion

Overall, the three studies highlight certain significant limitations in the original TWKSAT designed by Stevens and Campion (1999), improving the metric characteristics of the test and developing a new version. Available research shows that the TWKSAT offers good predictive validity (McClough & Rogelberg, 2003), but it may be improved in terms of reliability (e.g., Athanasaw, 2003; Chen et al., 2004). Despite the TWKSAT's relevance as an appropriate measure of essential teamwork competencies, the absence of studies examining the causes of these reliability problems is surprising.

Based on the results of Study 1, (a) the reliability problems in the TWKSAT are due to the

poor functioning of certain items, (b) the domain contents proposed by the authors are not satisfactorily represented by the test items, and (c) the dimensionality of the test does not reflect the original substantive model.

Study 2 presents the TWCT, which is a new development focused on the Stevens and Campion (1999) model. The TWCT includes 36 items in a 4-point frequency scale format and drafted in "observable behaviors" statements. The dimensionality of the TWCT reproduces the original substantive model better than its predecessor, reflecting the assessment contents domain included in the model and offering reliable scores in both the test total and each of the model's five dimensions.

Finally, Study 3 provides initial evidence for the criterion validity of the TWCT compared with the original test in a sample of professionals. The total TWCT score is positively correlated with the teamwork assessments made by both supervisors and the employees themselves. Moreover, the TWCT predicts the supervisor assessment better than the TWKSAT.

Overall, our research provides empirical evidence concerning the method for valid, reliable teamwork competency assessments. The TWCT offers a more accurate measure of teamwork competency in terms of reliability than the original measure proposed by Stevens and Campion (1994), as well as greater conceptual richness in terms of independent scores for the five competencies identified by the authors. This importantly allows differential analysis of the effects of specific competencies on the performance of different types of teams. In addition, it makes possible the examination of whether improving the skills associated with a specific teamwork competency benefits the others. Our findings also help both academics and practitioners to better understand the low reliability indicators for the TWKSAT and provide a new test (TWCT) that improves reliability and offers an independent measure for each of the dimensions proposed by Stevens and Campion (1999).

Considering the competencies measured by the TWCT from the standpoint of the five major teamwork dimensions proposed by Salas et al. (2005), the competencies measured at the individual level are found to be associated with two of the core dimensions these authors describe, namely follow-up and monitoring of team performance and team orientation. However, other

core dimensions of the model, such as team leadership, replacement behaviors, and adaptability, do not appear in the TWCT. In addition, recent studies have shown the capacity of knowledge tests on team roles related to the task, team, and boundary-spanning activities to accurately predict performance in teams (Marrone et al., 2007; Mumford et al., 2008). Future research should consider this issue as the TWCT would gain in both quality and utility if it included the mentioned dimensions.

Despite the above-mentioned contributions, our research is not without limitations. First, the samples analyzed in the first two studies are of a reasonable size but were provided by university students. If our results are to be generalized, the studies would need to be replicated in larger samples of employees. In particular, the factor analysis conducted on the TWKSAT should be replicated both in larger samples, given that our sample was smaller than the criterion of 10 participants per item, and in American samples to solve the potential effects of cultural differences. As proposed by Ryan, Chan, Ployhart, and Slade (1999), beyond the translation of the TWKSAT, it is necessary to adapt the test considering the equivalence of the measurements obtained in culturally diverse populations. Second, the factor analyses carried out on the TWCT reveal a dimensional structure that is similar to the substantive model employed but does not reflect it perfectly. Further studies should look into the reasons for this, particularly with regard to the functioning of the Collaborative Problem Solving scale. In line with the underlying theoretical model, it was expected that this competency would correlate with Conflict Management and Communication, but only a weak association was found. Moreover, Collaborative Problem Solving correlates more closely with other competencies related to team self-management. Analysis of the content of items (e.g., Item 11, "To address the trivial task-related issues, I do not need to talk first with all team members in order to reach a decision") indicates that behaviors of this kind are related to effective coordination.

The number and type of competencies selected are critical issues. The competencies originally proposed by Stevens and Campion (1994) do not exhaust the set of skills that could be considered to explain what people actually do in effective teamwork. For example, the

skills to develop shared mental models (Cannon-Bowers et al., 1995; Zaccaro, Rittman, & Marks, 2001), mutual trust (Webber, 2002), and team leadership (Cannon-Bowers et al., 1995; Marks, Mathieu, & Zaccaro, 2001) may also be critical in explaining team performance. As mentioned above, the development of measures that integrate these aspects would improve our ability to diagnose and predict teams' performance based on the personal abilities of the individuals who form them. Moreover, the use of advanced psychometric models like those based on item response theory (Hambleton & Swaminathan, 1985) would allow the design of computerized adaptive tests able to optimize the administration of a wide-ranging measure (Lord, 1970; Owen, 1975).

Finally, let us consider the practical implications of this research. The information provided by the new TWCT is of great interest for people management in organizations, as it shows managers, team leaders, and human resources professionals which employees' competencies are critical to select, train, and develop to improve team performance.

By throwing light on candidates' skills, the test would facilitate selection and team composition decisions. The access to information on five different competencies means that the configuration of teams can be supported by a richer information base than would be possible using only a single general estimate. In addition, the differential diagnosis provided by the TWCT on the stronger and weaker competencies of each employee will allow teams to customize training and development interventions. This is particularly valuable in view of the increasing demand for training programs to improve teamwork competencies (Chen et al., 2004). Finally, the behaviors made explicit by the TWCT could be used as a guide to determine the behavioral anchors and sample behaviors used in the design of performance assessment and management tools.

## References

- Athanasaw, Y. A. (2003). Team characteristics and team member knowledge, skills, and ability relationships to the effectiveness of cross-functional teams in the public sector. *International Journal of Public Administration*, 26, 1165–1203. doi: 10.1081/PAD-120019926

- Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology, 1*, 185–216. doi:10.1177/135910457000100301
- Cannon-Bowers, J. A., & Bowers, C. A. (2011). Team development and functioning. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology* (Vol. 1, pp. 597–650). Washington, DC: American Psychological Association. doi:10.1037/12169-019
- Cannon-Bowers, J. A., Tannenbaum, S. I., Salas, E., & Volpe, C. E. (1995). Defining competencies and establishing team training requirements. In R. A. Guzzo & E. Salas (Eds.), *Team effectiveness and decision making in organizations* (pp. 333–380). San Francisco, CA: Jossey-Bass.
- Chen, G., Donahue, L. M., & Klimoski, R. J. (2004). Training undergraduates to work in organizational teams. *Academy of Management Learning and Education, 3*, 27–40. doi:10.5465/AMLE.2004.12436817
- Clevenger, J., Pereira, G. M., Wiechmann, D., Schmitt, N., & Harvey, V. S. (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology, 86*, 410–417. doi:10.1037/0021-9010.86.3.410
- Cox, T., Lobel, S., & McLeod, P. (1991). Effects of ethnic group cultural differences on cooperative and competitive behavior on a group task. *Academy of Management Journal, 34*, 827–847. doi:10.2307/256391
- Earley, P. C., & Erez, M. (1997). *The transplanted executive*. New York, NY: Oxford University Press.
- Gibson, C. B. (1999). Do they do what they believe they can? Group efficacy and group effectiveness across tasks and cultures. *Academy of Management Journal, 42*, 138–152. doi:10.2307/257089
- Grimm, S. D., Church, A. T., Katigbak, M. S., & Reyes, J. A. S. (1999). Self-described traits, values, and moods associated with individualism and collectivism. *Journal of Cross-Cultural Psychology, 30*, 466–500. doi:10.1177/0022022199030004005
- Hackman, J. R. (2002). *Leading teams: Setting the stage for great performances*. Boston, MA: Harvard Business School Press.
- Hambleton, R. K., & Swaminathan, J. (1985). *Item response theory: Principles and applications*. Boston, MA: Kluwer.
- Hampden-Turner, C., & Trompenaars, F. (1993). *The seven cultures of capitalism: Value systems for creating wealth in the United States, Japan, Germany, France, Sweden, and the Netherlands*. Garden City, NY: Doubleday.
- Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*. Beverly Hills, CA: Sage.
- Hofstede, G. (1983a). The cultural relativity of organizational practices and theories. *Journal of International Business Studies, 14*, 75–89. doi:10.1057/palgrave.jibs.8490867
- Hofstede, G. (1983b). Dimensions of national cultures in fifty countries and three regions. In J. B. Deregowski, S. Dziurawiec, & R. C. Annis (Eds.), *Explications in cross-cultural psychology* (pp. 335–355). Lisse, Netherlands: Swets & Zeitlinger.
- Kluckhohn, F., & Strodtbeck, F. L. (1961). *Variations in value orientations*. Westport, CT: Greenwood Press.
- Kozlowski, S. W. J., & Ilgen, D. R. (2006). Enhancing the effectiveness of work groups and teams. *Psychological Science in the Public Interest, 7*, 77–124.
- Lawler, E. E. III, & Worley, C. G. (2006). *Built to change. How to achieve sustained organizational effectiveness*. San Francisco, CA: Jossey-Bass.
- Leach, D. J., Wall, T. D., Rogelberg, S. G., & Jackson, P. R. (2005). Team autonomy, performance, and member job strain: Uncovering the Teamwork KSA link. *Applied Psychology: An International Review, 54*, 1–24. doi:10.1111/j.1464-0597.2005.00193.x
- Lord, F. M. (1970). Some test theory for tailored testing. In W. H. Holtzman (Ed.), *Computer assisted instruction, testing and guidance* (pp. 139–183). New York: Harper & Row.
- Lorenzo-Seva, U., & Ferrando, P. J. (2006). FACTOR: A computer program to fit the exploratory factor analysis model. *Behavioral Research Methods, 38*, 88–91. doi:10.3758/BF03192753
- 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.
- Marrone, J. A., Tesluk, P. E., & Carson, J. B. (2007). A multi-level investigation of antecedents and consequences of team member boundary spanning behavior. *Academy of Management Journal, 50*, 1423–1439. doi:10.5465/AMJ.2007.28225967
- Maznevski, M. L., DiStefano, J. J., Gomez, C. B., Noorderhaven, N. G., & Wu, P. (1997, October 8–12). *The cultural orientations framework and international management research*. Paper presented at the Academy of International Business annual meeting, Guadalajara, Mexico.
- McClough, A. C., & Rogelberg, S. G. (2003). Selection in teams: An exploration of the Teamwork Knowledge, Skills and Ability Test. *International Journal of Selection and Assessment, 11*, 56–66. doi:10.1111/1468-2389.00226
- McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgement tests to predict job performance: A clarification of the literature. *Journal*

- of *Applied Psychology*, 86, 730–740. doi:10.1037/0021-9010.86.4.730
- Mumford, T. V., Van Iddekinge, C. H., Morgeson, F. P., & Campion, M. A. (2008). The Team Role Test: Development and validation of a team role knowledge situational judgment test. *Journal of Applied Psychology*, 93, 250–267. doi:10.1037/0021-9010.93.2.250
- Muthén, B., & Muthén, L. K. (2006). Mplus [Computer software]. Los Angeles, CA: Author.
- Owen, R. J. (1975). A Bayesian sequential procedure for quantal response in the context of adaptive mental testing. *Journal of the American Statistical Association*, 70, 351–356. doi:10.1080/01621459.1975.10479871
- Peeters, M. A. G., van Tuijl, H. F. J. M., Rutte, C. G., & Reymen, I. M. M. J. (2006). Personality and team performance: A meta-analysis. *European Journal of Personality*, 20, 377–396. doi:10.1002/per.588
- Rovinelli, R. J., & Hambleton, R. K. (1977). On the use of content specialists in the assessment of criterion-referenced test item validity. *Dutch Journal of Educational Research*, 2, 49–60.
- Ryan, A. M., Chan, D., Ployhart, R. E., & Slade, L. A. (1999). Employee attitude surveys in a multinational organization: Considering language and culture in assessing measurement equivalence. *Personnel Psychology*, 52, 37–58. doi:10.1111/j.1744-6570.1999.tb01812.x
- Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there a “big five” in teamwork? *Small Group Research*, 36, 555–599. doi:10.1177/1046496405277134
- Salas, E., Sims, D. E., & Klein, C. (2004). Cooperation at work. In C. D. Spielberger (Ed.), *Encyclopedia of applied psychology* (Vol. 1, pp. 497–505). Boston, MA: Elsevier Academic Press. doi:10.1016/B0-12-657410-3/00533-X
- Schmidt, F. L., & Hunter, J. E. (1983). Individual differences in productivity: An empirical test of estimates derived from studies of selection procedure utility. *Journal of Applied Psychology*, 68, 407–414. doi:10.1037/0021-9010.68.3.407
- Spencer, L. M., & Spencer, S. M. (1993). *Competence at work: Models for superior performance*. New York, NY: Wiley.
- Stevens, M. A., & Campion, M. J. (1994). The knowledge, skill, and ability requirements for teamwork: Implications for human resource management. *Journal of Management*, 20, 503–530.
- Stevens, M. A., & Campion, M. J. (1999). Staffing work teams: Development and validation of a selection test for teamwork settings. *Journal of Management*, 25, 207–228. doi:10.1016/S0149-2063(99)80010-5
- Thorn, D. W., & Deitz, J. C. (1989). Examining content validity through the use of content experts. *Occupational Therapy Journal of Research*, 9, 334–346.
- Triandis, H. C. (1995). *Individualism and collectivism*. Boulder, CO: Westview Press.
- Trompenaars, F. (1993). *Riding the waves of culture: Understanding cultural diversity in business*. Burr Ridge, IL: Irwin.
- Webber, S. S. (2002). Leadership and trust facilitating cross-functional team success. *Journal of Management Development*, 21, 201–214. doi:10.1108/02621710210420273
- Zaccaro, S. J., Rittman, A. L., & Marks, M. A. (2001). Team leadership. *The Leadership Quarterly*, 12, 451–483. doi:10.1016/S1048-9843(01)00093-5

(Appendices follow)

## Appendix A

### Observation Code Categories

#### Conflict Resolution

- She/he negotiates considering reasons instead of positions.
- She/he collects information about the different sources of conflict.
- She/he uses an integrative (win-win) negotiation strategy when possible.

#### Collaborative Problem Solving

- She/he describes problems in terms of behaviors, consequences, and feelings.
- She/he avoids setting conclusions and ascribing motives to others quickly.
- She/he focused on communalities to generate solutions (e.g., use brainstorming to generate more options).
- She/he requests additional information.

#### Communication

- She/he produces complete and specific messages.
- She/he asks for feedback about message understanding.

- She/he makes summaries along a communication interaction to demonstrate understanding.
- She/he communicates openly and supportively so that he/she recognizes the others.

#### Goal Setting and Performance Management

- She/he tries to define clearly the wished results (SMART objectives).
- She/he performs actions that promote others' success instead of hindering their progress.
- She/he shows awareness about others' performance (e.g., asking for participation of the most active members).

#### Planning and Task Coordination

- She/he makes questions for planning (e.g., How are we going to start the project? What resources are necessary? How much time we have to invest?).
- She/he clarifies the procedures of execution (what we make the first, the second, etc.).
- She/he helps to associate tasks with teammates.

## Appendix B

### TWCT Items

1. When my work team is in conflict, I try to make it explicit to find solution pathways.
2. When I interact with my team mates, I ask questions to better understand what they say.
3. When I disagree with others, I make an effort to focus on what we have in common instead of centering on what separates us.
4. I plan my tasks effectively.
5. I try to use the most appropriate team network to communicate the different types of information, avoiding the same formal procedure all the time.
6. I often get involved in monitoring the task performance of other team members.

*(Appendices continue)*

7. When we face an internal conflict because of a communication problem or misunderstanding, I try to solve it by asking questions and listening to the people involved.
8. I look at people when they talk to me and I modify my body language to show real interest in what they tell me.
9. I can easily recognize people's emotional states by observing their nonverbal messages.
10. If someone in my team acts inappropriately, I talk privately with her/him, encouraging the rest of the team to do the same.
11. To address the trivial task-related issues, I do not need to talk first with all team members so we reach a decision.
12. I make an effort to talk about less important things with my peers for the sake of team spirit and better internal communication.
13. Having knowledge about people's skills and situation requirements is critical to assign tasks properly.
14. Discussions without directions or guides can lead group members to make decisions that they would not make on their own.
15. When my personal interests are in conflict with others' interests, I tend to be honest in the negotiation so that others understand my needs.
16. I care and act to make team conflicts explicit in a way that they can be solved.
17. I play an active role in team meetings by offering my opinions, asking questions, and expressing my thoughts and ideas in a sincere and open way.
18. I often help others in my team to make clear the roles and tasks they have to perform.
19. When I am upset about something, I express my discomfort to the group in a constructive way, asking for solution alternatives.
20. I like to provide my peers with feedback about what they do and to assess and value their work.
21. If something upsets me in my team, I do not like to act as if nothing has happened.
22. I try to establish milestones in my work team so that we can monitor our assigned tasks.
23. When I am involved in a team project, I care about having clear plans concerning the tasks and the timing to accomplish them.
24. During group meetings, regulation is necessary to ensure that all members provide their opinions and to avoid that only a few participate actively.
25. When performing tasks in which one is an expert, the contributions made by other members are not that important.
26. In group decision meetings, it is more usual to promote cohesion and reach a majority agreement than to pay attention to divergent opinions.
27. I try listening to my peers' opinions without evaluating their positions as good or bad.
28. When working in a group, I say what I think in an open and sincere way.
29. I expect my peers trust enough to tell me the aspects of my work that they most dislike.
30. I sometimes talk with my peers without an objective, just for sharing a while together.
31. It is important for me to monitor the tasks assigned to each team member.
32. I provide my peers with relevant information on how well I think the team tasks are progressing.
33. When doing my job, I prioritize the tasks most necessary for my teammates to complete their work.
34. I try to ensure that my outputs match the inputs needed by my peers to perform their tasks.
35. For the sake of team work, I set objectives with moderate difficulty so that effort is needed to accomplish them.
36. I often provide my peers with feedback on their task performance.

Received October 15, 2012

Revision received November 7, 2013

Accepted December 14, 2013 ■

## The Effect of Different Phases of Synchrony on Pain Threshold

Philip J. Sullivan, Kate Rickers, and Kimberley L. Gammage  
Brock University

Synchronization of behavior between individuals has been found to result in a variety of prosocial outcomes. The role of endorphins in vigorous synchronous activities (Cohen, Ejsmond-Frey, Knight, & Dunbar, 2010) may underlie these effects as endorphins have been implicated in social bonding (Dunbar & Shultz, 2010). Although research on synchronous behavior has noted that there are 2 dominant phases of synchrony—in-phase and antiphase synchrony (Marsh, Richardson, Baron, & Schmidt, 2006), research on the effect of synchrony on endorphins has only incorporated in-phase synchrony. The current study examined whether both phases of synchrony would generate the synchrony effect. Twenty-two participants rowed under 3 counter-balanced conditions—alone, in-phase synchrony and antiphase synchrony. Endorphin release, as measured by pain threshold, was assessed before and after each session. Change in pain threshold during the in-phase synchrony session was significantly higher than either of the other 2 conditions. These results suggest that the synchrony effect may be specific to just in-phase synchrony, and that social presence is not a viable explanation for the effect of synchrony on pain threshold.

*Keywords:* behavioral synchrony, endorphins, pain threshold, in-phase synchrony

People tend to be inclined to move synchronously, whether to an external stimuli or to partners. For example, children as young as two have been found to impulsively synchronize themselves with a drum beat (Kirschner & Tomasello, 2009), and even minimal groups, such as two individuals in rocking chairs, will spontaneously synchronize their motions (Richardson, Marsh, Isenhower, Goodman, & Schmidt, 2007). The relevance of synchronized behavior may be seen in a variety of interpersonal outcomes that have resulted from joint action or mimicry. Whether people spontaneously act in coordination with each other, or one person mimics the actions of someone else, there appears to be a robust effect of such actions on interpersonal attraction and prosocial behavior.

Acting in synchrony with others has repeatedly been found to increase perceptions of interpersonal attraction between participants. Peo-

ple have been shown to infer greater cohesiveness on individuals who move in synchrony compared with those who move independently (Lakens & Stel, 2011). Synchronization of individuals' movements with others has been shown to result in higher ratings of interpersonal attraction and affiliation (Bernieri, 1988; LaFrance & Broadbent, 1976; Miles, Nind, & Macrae, 2009).

Hove and Risen (2009) conducted a series of experiments on interpersonal attraction between individuals engaged in synchronized behavior. In one study, participants engaged in synchronized finger tapping with an experimenter for a 2.5 min trial. Degree of synchrony, as measured by percentage of finger taps occurring within 100 ms, was significantly correlated with higher ratings of liking for the experimenter. Their second experiment manipulated interpersonal synchrony and included a baseline measure of liking. Results showed that individuals randomly assigned to a condition in which they tapped in synchrony with the experimenter reported significantly greater liking for the experimenter than those in either the asynchrony or alone conditions. There were no significant baseline differences in liking for the experimenter.

---

Philip J. Sullivan, Kate Rickers, and Kimberley L. Gammage, Department of Kinesiology, Brock University.

Correspondence concerning this article should be addressed to Philip J. Sullivan, Department of Kinesiology, Brock University, St. Catharines, ON, L2S 3A1. E-mail: [phil.sullivan@brocku.ca](mailto:phil.sullivan@brocku.ca)

Synchrony has also been repeatedly linked to a variety of prosocial behaviors. Individuals who act in synchrony tend to display significantly greater cooperation (Valdesolo, Ouyang & DeSteno, 2010; Wiltermuth & Heath, 2009), norm compliance (Wiltermuth, 2012), compassion (Valdesolo & DeSteno, 2011), and generosity (Stel, van Baaren, & Vonk, 2007) than individuals who do not act in synchrony.

Wiltermuth and Heath (2009) investigated the effect of synchronized behavior and cooperation in small groups. Participants walked around campus in groups of three in one of two conditions. In the synchronized condition, they walked in step; in the nonsynchronized condition, they walked normally. In a postwalk cooperative game, individuals in the synchronized condition behaved significantly more cooperatively than those in the nonsynchronized condition. A second experiment in the same study replicated this effect, even after controlling for common identity and common fate within the groups.

In a study conducted by Valdesolo, Ouyang, and DeSteno (2010), individuals were randomly assigned to one of two conditions. They rocked in rocking chairs in pairs in either a synchronized or unsynchronized fashion for two 90 s trials. After this manipulation, the pairs had to participate in a cooperative joint action task. Subjects also completed ratings of similarity, connectedness, and liking for their partners. Pairs in the synchronized condition completed the cooperative task in significantly less time than those in the unsynchronized condition. Synchrony also resulted in higher perceptions of connectedness and liking.

Recent research may have uncovered a potential mechanism for these social effects of synchrony. It appears that synchronous behaviors, particularly vigorous synchronous behaviors, elevate participants' endorphin levels, which have been linked to social bonding in humans and primates (Dunbar, 2010; Dunbar & Shultz, 2010). It has been suggested that endogenous endorphins may mediate between affiliative stimuli (e.g., shared synchronous movements) and affiliation (DePue & Morrone-Strupinsky, 2005). Dunbar and Shultz (2010) stated that social grooming in primates builds intense social bonds through the release of endorphins, "which provides a psycho-pharmacological mechanism that enables two individuals to build a bonded

relationship with some kind of deeply emotional basis" (p. 782).

Because brain endorphins do not cross the blood-brain barrier, and can only be measured through an invasive lumbar puncture (Boeker et al., 2008; Dearman & Francis, 1983), pain threshold has been commonly used as an indicator of endorphin activity. (Dunbar et al., 2012; Jamner & Leigh, 1999; Zillman, Rockwell, Schweitzer, & Sundar, 1993). Pain threshold is understood to be a valid indicator of endorphin activity because endogenous opioids are released during pain sensation (DePue & Morrone-Strupinsky, 2005; Mueller, 1981). Specifically, the use of a blood pressure cuff to induce specific ischemic pain has been found to be a valid and reliable, as well as noninvasive, protocol to assess endorphin levels (Estebe, Le Naoures, Chemaly, & Ecoffey, 2000; Ryan & Kovacic, 1966).

Cohen, Ejsmond-Frey, Knight, and Dunbar (2010) investigated synchronized behaviors in a team of male collegiate rowers, using the pain threshold test as a noninvasive measure of central endorphin activity. They found a difference when teammates rowed in synchronicity in a virtual boat compared with when they performed the same workout alone. Specifically, the synchronized condition produced a significantly greater increase in pain threshold from pre- to postexercise than the solitary session. This was despite the fact that workload was standardized between the two conditions. Cohen et al. (2010) concluded that vigorous synchronized activity results in a "synchrony effect" whereby endorphin activity is increased.

This synchrony effect has been replicated by Sullivan and Rickers (2013) who used the same protocol as Cohen et al. (2010), with participants who were strangers as well as rowing teammates. Because of the role of endorphins in social bonding, the study investigated whether the use of teammates in Cohen's study may have influenced the results. Sullivan and Rickers found that when individuals rowed for 45 min with either teammates or strangers, they still displayed significantly higher changes in pain threshold than when they rowed alone.

One aspect of synchrony that may be relevant to the synchrony effect is the phase of synchrony. There are two dominant phases of synchrony—in-phase and antiphase (Marsh, Richardson, Baron, & Schmidt, 2006). In-phase



synchrony is characterized by movements being in perfect unison whereas antiphase symmetry is characterized by alternating patterns of movement between two individuals. For example, if two children were side by side on swings, in-phase movement would be if they both swing forward and then backward together. Antiphase synchrony would be seen when one child swings forward as the other swings back (Miles, Nind, Henderson, & Macrae, 2010). Humans appear to prefer in-phase synchrony (Marsh et al., 2006; Richardson et al., 2007). Furthermore, in-phase synchrony appears to have a more powerful effect on some of the outcomes noted above (Macrae, Duffy, Miles, & Lawrence, 2008).

Although these two different phases may both affect outcomes of synchronized behavior, to date the synchrony effect has only been investigated with in-phase synchrony. The current study investigated the role of different phases of behavioral synchrony on the synchrony effect. Specifically, rowing under solitary, in-phase and antiphase synchrony conditions were examined to see if all would produce the same effect on pain threshold that was seen by Cohen et al. (2010) and Sullivan and Rickers (2013). Because previous research has noted significantly greater increases in endorphin levels after synchronous compared with solitary activity, it was predicted that there would be a significant difference between solitary and in-phase synchrony. Because antiphase synchrony has not yet been studied with respect to the synchrony effect, no specific hypotheses were put forward regarding this condition.

## Method

### Participants

Twenty-two (9 male, 13 female) individuals participated in the study. Participants ranged in age from 19 to 55 years ( $M = 31.14$ ,  $SD = 13.13$ ). Men's average height was 73.11 in. ( $SD = 1.54$ ) and weight was 200.56 pounds ( $SD = 23.91$ ). Women's average height was 67.62 in. ( $SD = 1.94$ ) and weight was 145.00 ( $SD = 12.75$ ). All participants had rowed for at least 1 year.

### Procedures

Before data collection, this study was approved by the IRB of the primary author. Before participation, all participants completed informed-consent forms and the Physical Activity Readiness Questionnaire (PAR-Q; Canadian Society for Exercise Physiology, 2002). The PAR-Q is a 7-item measure to assess whether participants can safely partake in physical activity. Every participant was required to answer to questions using either yes or no responses. If they replied with "yes" to any of the questions, they were to obtain permission from their doctors to be physically active. If they answered "no" to all questions, it was safe for them to participate in physical activity. All participants were able to partake in the procedure.

The experiment used a randomized counterbalanced repeated measures design. Participants rowed in all of three conditions—alone, and in two synchronized condition with a partner. In the in-phase synchrony condition, both rowers kept in the same movement pattern (i.e., both rowers would be at the fully extended and full contracted positions of the stroke at the same time). In the antiphase synchrony condition, rowers kept at opposite points of the movement pattern (i.e., when one rower was fully extended, the other rower would be fully contracted, and vice versa). These conditions were counterbalanced so that one third of the participants rowed alone first, one third rowed in the in-phase synchrony condition first, and the remaining third rowed in the antiphase synchrony condition first.

Intensity of the workout was operationalized by average 500 m split time (i.e., the time taken to row 500 m on the ergometer machine) over the 30 min session. This value was recorded during the first session and participants matched the intensity on subsequent trials. By slight manipulations of the drag factor, the experimenters were able to establish synchrony between participants while allowing individuals to maintain the same intensity (i.e., stroke rate) in all three conditions.

Participants rowed with the same partner in both synchrony conditions and pairs always consisted of rowers of the same gender. All participants were prohibited from listening to music or consuming sport drinks during their trials. There was a week gap between sessions

for all participants. All sessions were held in the afternoon or early evening.

Measures of pain threshold were taken before and after all trials using the blood pressure cuff protocol used by Cohen et al. (2010). All pain measurements were taken by the experimenter and recorded ~1 min before the start and conclusion of each trial. In all conditions, the same experimenter assessed pain threshold for all participants. This was achieved by staggering the end of workouts of participants by ~30 s each. The pain thresholds of all participants were assessed away from all others being tested. The blood pressure cuff was placed above the elbow on the nondominant arm of each participant. Ischemic pain was induced through manual pumping of the cuff to increase pressure. Participants expressed the point at which pressure became uncomfortable by saying “now,” at which point the pressure was recorded and the cuff was removed. This protocol has been commonly used in studies on behavioral synchrony (Cohen et al., 2010; Dunbar et al., 2012; Sullivan & Rickers, 2013), and previous research has supported its validity as a measure of pain threshold. It has been shown to be highly correlated with other measures of pain threshold, including visual analog scales (Estebe et al., 2000) and gross pressure ischemia (Ryan & Kovacic, 1966). Furthermore, it has been found to have a high test–retest reliability within participants ( $r = .87$ ; Ryan & Kovacic, 1966).

## Materials

Two Concept2 Model D indoor rowing machines (i.e., “ergometers”) were set up in the rowing training center of the university. Pain threshold was measured using an AMG Medical Inc. Professional Series sphygmomanometer (blood pressure cuff). Units of pressure were recorded in millimeters of mercury (mmHg).

## Results

Because pain threshold differs by individual (Dunbar et al., 2012), change in pain threshold from pre- to postscore was used as the dependent variable in analyses. Although there are some concerns with using raw change scores in such analyses, research has shown that when using reliable scores, particularly physiological measures, it is advisable to use raw change

scores (Dimitrov & Rumrill, 2003). Pre- and postpain threshold scores by condition are shown in Table 1. All three conditions displayed mean increases in pain threshold from pre- to postactivity, which is consistent with the effect of vigorous exercise on endorphin activity (Boecker et al., 2008). Change in pain threshold was normally distributed in all three conditions. Average prescores did not differ between conditions ( $F(2, 63) = 0.89, p = .416$ ), and there were significant pre- to postchanges in the solitary ( $F(1, 21) = 24.10, p < .001$ ), in-phase ( $F(1, 21) = 69.69, p < .001$ ) and antiphase conditions ( $F(1, 21) = 22.04, p < .001$ ). There were no statistically significant differences between genders in preactivity pain threshold scores or changes scores in any of the three conditions.

Before testing the hypotheses, the data was examined to see if the assumption of sphericity was upheld. Mauchly’s test was nonsignificant ( $W = .929, p > .05$ ), indicating that there was homogeneous variance between the conditions, and that the data was appropriate for a repeated measures analysis. A repeated measures analysis of variance (ANOVA) revealed a significant effect for condition ( $F(2, 42) = 4.12, p = .023, \eta^2 = .16$ ). Observed power within the design was .70. Bonferroni adjusted contrast analyses revealed that the in-phase synchrony condition was significantly different from both the solitary ( $F(1, 21) = 5.27, p = .03, \eta^2 = .20$ ) and antiphase ( $F(1, 21) = 8.99, p = .007, \eta^2 = .30$ ) conditions. There was not a significant difference between the solitary and antiphase conditions ( $F(1, 21) = 0.41, p > .05$ ).

## Discussion

The current study was designed to investigate if two different phases of synchrony would both

Table 1  
Means and SDs of Raw and Change Pain  
Threshold Scores by Condition

Condition	Prescore ( <i>M, SD</i> )	Postscore ( <i>M, SD</i> )	Change score ( <i>M, SD</i> )
Solitary	178.64 (38.95)	210.00 (51.73)	31.36 (29.96)
In-phase	196.82 (55.41)	247.73 (50.89)	50.91 (28.60)
Antiphase	194.55 (51.89)	224.09 (52.34)	29.55 (29.52)

Note. Values are in millimeters of mercury (mmHg).

show the “synchrony effect” of elevated pain threshold, which has previously only been documented with in-phase synchrony (Cohen et al., 2010; Sullivan & Rickers, 2013). Within the current study, it was found that when individuals rowed in an in-phase synchronous condition, they reported a significantly greater increase (from pre- to postactivity) in pain threshold than either the solitary or antiphase synchrony conditions. There was no significant difference between the antiphase and solitary condition. Therefore, the synchrony effect appears to be only a result of in-phase synchrony.

At this point, it appears that the synchrony effect is a fairly robust phenomenon. It has repeatedly been found that individuals have a significantly higher change in pain threshold after vigorous activity in synchrony with others than after the same vigorous activity performed alone (Cohen et al., 2010; Sullivan & Rickers, 2013). This has been found with different samples, including teammates and strangers, and both males and females. Whereas Cohen et al.’s (2010) original study only included male rowers who were teammates, Sullivan and Rickers (2013) included strangers as well as teammates and males as well as females, and the current study also found the effect using both male and female participants.

As pain threshold has been interpreted as an indication of endorphin activity (Cohen et al., 2010; Dunbar et al., 2012), it appears that vigorous synchronized activity may produce a significant “rush” in endorphins compared with the same activity performed alone. Endorphins have been implicated in social bonding, although the exact mechanism is not clearly understood. As noted above, it is possible that endorphins mediate affiliative stimuli, such as synchronous movements, and affiliation (DePue & Morrone-Strupinsky, 2005). If this is the case, it may be that synchronized movements, particularly in-phase synchrony, which appears to be preferred by people (Marsh et al., 2006; Richardson et al., 2007), causes endorphin release. In the case of vigorous synchronized activities, such as rowing, this endorphin release is above and beyond what is induced by the exercise itself (Cohen et al., 2010). Because one of the effects of endorphins, in addition to pain threshold, is feelings of well-being (Dunbar & Shultz, 2010), this shared synchrony would result in stronger social bonds among participants.

Therefore, it is possible that through the synchrony effect, vigorous synchronized activity releases endorphins, which in turn increase social bonding. This process would be consistent with speculation that collective activities increase cohesion among group members in activities such as marching in step and rhythmic dancing and singing during religious activities (McNeill, 1995). Through the inducement of endorphins, the synchrony effect may serve as a mechanism for this increased cohesion.

A secondary implication of the finding that in-phase synchrony produces an effect on endorphins whereas antiphase synchrony does not is that this synchrony effect now appears to be independent of any social facilitation effect. The designs of Cohen et al., (2010) and Sullivan and Rickers (2013) both included a confounding factor—the synchrony condition was also the only group condition. In tasks requiring physical conditioning demands, such as rowing, the presence of others should result in performance increases (Strauss, 2002). Although performance was constant in both conditions in Cohen et al.’s and Sullivan and Ricker’s research, it may be that other responses were affected. The presence of others has been known to have an effect on unlearned responses, particularly cardiovascular activity. Blascovich, Mendes, Hunter, and Salomon (1999) found that the presence of others during well-learned tasks could produce a “challenge” response (i.e., an increase in cardiovascular activity). Because exercise stimulates endorphin release (Boeker et al., 2008), such a change could be responsible for the change in endorphins seen by Cohen and her colleagues. It may have been possible that the endorphin effect that Cohen attributed to synchrony was actually a mere presence effect. However, given the current results, where a second group condition (i.e., the antiphase synchrony) did not produce this synchrony effect, it appears that the effect on endorphins is because of synchrony, in particular in-phase synchrony, rather than any social facilitation effect.

The present results extend our understanding of the synchrony effect. Consistent with previous research, they have supported the finding that vigorous synchronized activity causes an increase in pain threshold, which appears to be an indication of elevated endorphin activity. This effect now appears to be the result of

synchrony, as opposed to social coaction, and specifically in-phase synchrony. Future research is still required to fully understand this synchrony effect. Specifically, the studies to date have only used the synchronized activity of rowing. Would the same synchrony effect be seen with other activities such as running in step, or fitness classes? It would also be interesting to see if the effect could be replicated with synchronized nonvigorous activities such as tai chi or walking in step. Furthermore, we would note that the links between synchrony and endorphins and cohesion are currently speculative and research is required to fully operationalize this potential mechanism. To fully understand this phenomenon, we need to know if synchrony directly effects endorphins and if so, does this synchrony-induced endorphin activity affect cohesion. As interesting as the effect appears to be, the time is ripe for more research on the basic mechanics and ramifications of this synchrony effect.

However, there are several limitations to the current design, which are probably not limited to the present study. The current sample, like Cohen et al.'s (2010) and Sullivan and Rickers (2013) comprised experienced rowers. It is possible that experienced rowers may associate synchrony with good teamwork and the resultant pain threshold effect may be related to this social effectiveness as opposed to any synchrony effect. Furthermore, there are some potential demand characteristics in the study. In this protocol, the experimenter is not blind to the condition of the participants, and it is possible that this may have influenced the results. Future research on the effect of behavioral synchrony on pain threshold or endorphin activity should use more strict methodologies, including blinded conditions and exit questionnaires to probe participants' suspicions about research questions.

## References

- Bernieri, F. (1988). Coordinated movement and rapport in teacher-student interactions. *Journal of Nonverbal Behavior*, *12*, 120–138. doi:10.1007/BF00986930
- Blascovich, J., Mendes, W. B., Hunter, S. B., & Salomon, K. (1999). Social “facilitation” as challenge and threat. *Journal of Personality and Social Psychology*, *77*, 68–77. doi:10.1037/0022-3514.77.1.68
- Boeker, H., Sprenger, T., Spilker, M. E., Henriksen, G., Koppenhoefer, M., Wagner, K. L., . . . Tolle, T. R. (2008). The runner's high: Opioidergic mechanisms in the human brain. *Cerebral Cortex*, *18*, 2523–2532. doi:10.1093/cercor/bhn013
- Canadian Society for Exercise Physiology. (2002). *Physical Activity Readiness Questionnaire (PAR-Q)*. Retrieved from <http://www.csep.ca/communities/c574/files/hidden/pdfs/par-q.pdf>
- Cohen, E. E., Ejsmond-Frey, R., Knight, N., & Dunbar, R. I. (2010). Rowers' high: Behavioural synchrony is correlated with elevated pain thresholds. *Biology Letters*, *6*, 106–108. doi:10.1098/rsbl.2009.0670
- Dearman, J., & Francis, K. T. (1983). Plasma levels of catecholamines, cortisol, and beta-endorphins in male athletes after running 26.2, 6, and 2 miles. *The Journal of Sports Medicine and Physical Fitness*, *23*, 30–38.
- DePue, R. A., & Morrone-Strupinsky, J. V. (2005). A neurobehavioral model of affiliative bonding: Implications for conceptualization a human trait of affiliation. *Behavioral Brain Science*, *28*, 313–385. doi:10.1017/S0140525X05290064
- Dimitrov, D. M., & Rumrill, Jr, P. D. (2003). Pretest-posttest designs and measurement of change. *Work: A Journal of Prevention, Assessment and Rehabilitation*, *20*, 159–165.
- Dunbar, R. I. M. (2010). The social role of touch in humans and primates: Behavioural function and neurobiological mechanisms. *Neuroscience and Biological Reviews*, *34*, 260–268. doi:10.1016/j.neubiorev.2008.07.001
- Dunbar, R. I. M., Baron, R., Frangou, A., Pearce, E., van Leeuwen, J. C., Stow, J., . . . van Vugt, M. (2012). Social laughter is correlated with an elevated pain threshold. *Proceedings of The Royal Society Biological Sciences*, *279*, 1161–1167. doi:10.1098/rspb.2011.1373
- Dunbar, R. I., & Shultz, S. (2010). Bondedness and sociality. *Behavior*, *147*, 775–803. doi:10.1163/000579510X501151
- Estebe, J. P., Le Naoures, A., Chemaly, L., & Ecofey, C. (2000). Tourniquet pain in a volunteer study: Effect of changes in cuff width and pressure. *Anaesthesia*, *55*, 21–26.
- Hove, J. M., & Risen, L. J. (2009). It's all in the timing: Interpersonal synchrony increases affiliation. *Social Cognition*, *27*, 949–960.
- Jamner, L. D., & Leigh, H. (1999). Repressive/defensive coping, endogenous opioids and health: How a life so perfect can make you sick. *Psychiatry Research*, *8*, 17–31. doi:10.1016/S0165-1781(98)00134-6
- Kirschner, S., & Tomasello, M. (2009). Joint drumming: Social context facilitates synchronization in preschool children. *Journal of Experimental Child Psychology*, *103*, 1–10. doi:10.1016/j.jecp.2009.06.001

- Psychology*, 102, 299–314. doi:10.1016/j.jecp.2008.07.005
- LaFrance, M., & Broadbent, M. (1976). Group rapport: Posture sharing as a nonverbal indicator. *Group and Organizational Management*, 1, 328–333. doi:10.1177/105960117600100307
- Lakens, D., & Stel, M. (2011). If they move in sync, they must feel in sync: Movement synchrony leads to attributions of rapport and entitativity. *Social Cognition*, 29, 1–14. doi:10.1521/soco.2011.29.1.1
- Macrae, C. N., Duffy, O. K., Miles, L. K., & Lawrence, J. (2008). A case of hand waving: Action synchrony and person perception. *Cognition*, 109, 152–156. doi:10.1016/j.cognition.2008.07.007
- Marsh, K. L., Richardson, M. J., Baron, R. M., & Schmidt, R. C. (2006). Contrasting approaches to perceiving and acting with others. *Ecological Psychology*, 18, 1–38. doi:10.1207/s15326969eco1801\_1
- McNeill, W. H. (1995). *Keeping together in time: Dance and drill in human history*. Cambridge, MA: Harvard University Press.
- Miles, L. K., Nind, L. K., Henderson, Z., & Macrae, C. N. (2010). Moving memories: Behavioral synchrony and memory for self and others. *Journal of Experimental Social Psychology*, 46, 457–460. doi:10.1016/j.jesp.2009.12.006
- Miles, L. K., Nind, L. K., & Macrae, C. N. (2009). The rhythm of rapport: Interpersonal synchrony and social perception. *Journal of Experimental Social Psychology*, 45, 585–589. doi:10.1016/j.jesp.2009.02.002
- Mueller, G. P. (1981). Beta-endorphin immunoreactivity in rat plasma: Variations in response to different physical stimuli. *Life Sciences*, 29, 1669–1674. doi:10.1016/0024-3205(81)90069-2
- Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R., & Schmidt, R. C. (2007). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science*, 26, 867–891. doi:10.1016/j.humov.2007.07.002
- Ryan, E. D., & Kovacic, C. R. (1966). Pain tolerance and athletic participation. *Perceptual and Motor Skills*, 22, 383–390. doi:10.2466/pms.1966.22.2.383
- Stel, M., van Baaren, R. B., & Vonk, R. (2007). Effects of mimicking: Acting prosocially and being emotionally moved. *European Journal of Social Psychology*, 38, 965–976. doi:10.1002/ejsp.472
- Strauss, B. (2002). Social facilitation in motor tasks: A review of theory and research. *Psychology of Sport and Exercise*, 3, 237–256. doi:10.1016/S1469-0292(01)00019-x
- Sullivan, P., & Rickers, K. (2013). The effects of behavioral synchrony in groups of teammates and strangers. *International Journal of Sport and Exercise Psychology*, 1–6. doi:10.1080/1612197X.2013.750139
- Valdesolo, P., & DeSteno, D. (2011). Synchrony and the social tuning of compassion. *Emotion*, 11, 262–266. doi:10.1037/a0021302
- Valdesolo, P., Ouyang, J., & DeSteno, D. (2010). The rhythm of joint action: Synchrony promotes cooperative ability. *Journal of Experimental Social Psychology*, 46, 693–695. doi:10.1016/j.jesp.2010.03.004
- Wiltermuth, S. (2012). Synchrony and destructive obedience. *Social Influence*, 7, 78–89. doi:10.1080/15534510.2012.658653
- Wiltermuth, S. S., & Heath, C. (2009). Synchrony and cooperation. *Psychological Science*, 20, 1–5. doi:10.1111/j.1467-9280.2008.02253.x
- Zillman, D., Rockwell, S., Schweitzer, K., & Sundar, S. (1993). Does humor facilitate coping with physical discomfort? *Motivation Emotion*, 17, 1–21. doi:10.1007/BF00995204

Received February 17, 2013

Revision received July 18, 2013

Accepted July 22, 2013 ■

## Reactions to Social Inclusion and Ostracism as a Function of Perceived In-Group Similarity

Donald F. Sacco

The University of Southern Mississippi

Michael J. Bernstein

Pennsylvania State University–Abington

Steven G. Young

Baruch College, City University of New York

Kurt Hugenberg

Miami University

Although ostracism is a powerfully aversive experience, recent evidence identifies factors capable of moderating the impact of ostracism, such as in-group status and the group's essential nature. In the current work, 67 Caucasian American participants (47 women) were included or ostracized by either same-race (i.e., Caucasian American) or other-race (i.e., African American) targets on a between-subjects basis while playing the game Cyberball. Participants then indicated the extent to which they felt similar to the other Cyberball players as well as how satisfied their basic needs (e.g., belongingness, self-esteem) were during the game. Consistent with past research, we found that in-group and out-group status moderated the magnitude of reactions to social inclusion and ostracism; that is, ostracism hurts more and social inclusion feels better when it is implemented by fellow in-group as opposed to out-group members. Importantly, we extend these previous findings by demonstrating that differential reactions to social inclusion and ostracism are mediated by changes in participants' self-perceived similarity with in-group members. These results identify a potential mechanism responsible for the differential impact of in-group–out-group status on reactions to social inclusion and ostracism.

*Keywords:* ostracism, intergroup relations, in-group–out-group, threat

Humans are intensely social animals. As such, we place great value on social connections and group memberships, with our self-esteem and emotional states being closely linked to our social interactions and relationships (Baumeister & Leary, 1995). As a result, being socially rejected or ostracized is an intensely aversive social experience, whereas being socially accepted is decidedly positive (MacDonald &

Leary, 2005; Williams, 2007). However, there has been some debate as to whether all experiences of ostracism and social inclusion are equally impactful. Although initial research indicated that aversive responses to ostracism are reflexive and unmoderated by situational factors (e.g., Gonsalkorale & Williams, 2007; Williams & Zadro, 2005), more recent research suggests that responses to social ostracism are sensitive to several factors, including characteristics of the individuals engaging in the act of ostracism. For example, Bernstein and colleagues (2010) demonstrated that being included or ostracized by in-group members exacerbates the intensity of the experience (i.e., being ostracized by an in-group hurts more than being ostracized by an out-group).

Whereas these past findings demonstrate that in-group status is an important moderator of reactions to social inclusion and ostracism, research has yet to identify the mechanism under-

---

Donald F. Sacco, Department of Psychology, The University of Southern Mississippi; Michael J. Bernstein, Department of Psychological and Social Sciences, Pennsylvania State University–Abington; Steven G. Young, Department of Psychology, Baruch College, City University of New York; Kurt Hugenberg, Department of Psychology, Miami University.

Correspondence concerning this article should be addressed to Donald F. Sacco, Department of Psychology, The University of Southern Mississippi, Owings-McQuagge Hall, 118 College Drive 5025, Hattiesburg, MS 39406. E-mail: donald.sacco@usm.edu

lying these differential reactions. To address this research lacuna, we tested the hypothesis that inclusion within and ostracism by an in-group can have more powerful influences on the self (e.g., influencing belongingness and self-esteem) precisely because such experiences can change perceptions of self-similarity to the in-group, which serves as a psychological mediator of the increased intensity of in-group inclusion and ostracism. To this end, we first review the literature on the need to belong, in-group-out-group status, and perceptions of similarity, and then provide direct experimental evidence for our proposed mediational hypothesis.

### The Need to Belong

Whereas stable social bonds have numerous physical and psychological benefits, unfulfilled belongingness needs have a decidedly negative impact on well-being (Baumeister & Leary, 1995). Specifically, acute experiences of ostracism induce negative mood and thwart basic social needs satisfaction (e.g., self-esteem, belonging; Williams & Zadro, 2005). Importantly, these strong reactions to the experience of ostracism are largely unaffected by various individual difference measures that seem intuitively relevant (e.g., trait self-esteem and social anxiety; Leary, Haupt, Strausser, & Chokel, 1998; Zadro, Boland, & Richardson, 2006).

Indeed, prior work has even found that ostracism lowers mood and basic need satisfaction equally, regardless of the characteristics of the ostracizing party. For example, Williams and colleagues (2000) found that responses to ostracism did not differ based on in-group-out-group distinctions (e.g., PC vs. Mac users; see also Smith & Williams, 2004). Furthermore, individuals' basic needs satisfaction is threatened regardless of whether participants believe that they were ostracized by fellow participants (i.e., peers), by confederates who were trained to ostracize them, or even by a computer programmed to arbitrarily ostracize them (Zadro, Williams, & Richardson, 2004). Finally, Gonsalkorale and Williams (2007) even found that ostracism perpetrated by a disliked out-group (i.e., KKK members) is experienced as painfully as that perpetrated by a valued in-group.

Nonetheless, not all sources of affiliation equally satisfy belonging needs. Specifically, valued social groups are a vital source of social

and material support, offering protection and access to mates (e.g., Correll & Park, 2005). Indeed, in-group bias, or the tendency to favor in-groups over out-groups (Tajfel, 1982), is reflected in numerous behaviors, ranging from more positive evaluations of in-groups, greater levels of prosocial behavior toward fellow in-group members, and allocation of more resources to in-groups (e.g., Ferguson & Kelley, 1964; Hein, Silani, Preuschoff, Batson, & Singer, 2010; Tajfel, Billig, Bundy, & Flament, 1971). Consistent with the importance of social in-groups, Bernstein and colleagues (2010) recently reported two studies demonstrating that ostracism is more painful and social inclusion more positive when carried out by fellow in-group members (cf. Goodwin, Williams, & Carter-Sowell, 2010). In their first study, participants were either ostracized or socially included by racial in-group or out-group members; participants reported more threatened affiliation needs following racial in-group ostracism (relatively to ostracism by racial out-group members), but more satisfied basic affiliation needs following social inclusion by racial in-group as opposed to out-group members. Their second study replicated this basic finding with political group membership; that is, participants reported more social pain when ostracized by individuals who shared their own party affiliation (as opposed to ostracism by confederates of a different party affiliation), but greater satisfaction when socially included by individuals with the same political party affiliation as their own. Importantly, these authors found that this strengthened experience of in-group inclusion and ostracism occurred most strongly for perceivers who were made to see the in-group as highly essentialized (i.e., participants who were led to believe that these groups possessed characteristics defined as relatively inborn, immutable, or genetically based; see Prentice & Miller, 2007).

### Perceived Self-Similarity and Reactions to In-Group-Out-Group Inclusion and Ostracism

In the current work, we tested the hypothesis that the experience of becoming psychologically similar to the including in-group or dissimilar from the ostracizing in-group would mediate the effect of in-group-out-group status on

reactions to inclusion and ostracism. Across numerous kinds of relationships, individuals prefer to interact with others who are similar across a variety of domains, including intelligence and physical attractiveness (e.g., Feingold, 1988; Jensen, 1978). Furthermore, similarity is positively related to relationship satisfaction and longevity in both platonic and romantic relationships (e.g., Acitelli, Kenny, & Weiner, 2001; Gonzaga, Campos, & Bradbury, 2007; Russell & Wells, 1991).

Interestingly, this relation between similarity and liking is bidirectional, such that positive interactions with others lead to increased perceptions of similarity. According to Morry's (2005) attraction-similarity model, attraction leads to perceptions of similarity between the self and others and these perceptions of similarity provide numerous benefits, ranging from feeling more understood and validated to increased feelings of positive affect and decreased loneliness. Put simply, we are likely to affiliate with similar others (e.g., Acitelli et al., 2001) but also perceive those we affiliate with as more similar to the self, and these feelings of interpersonal similarity are important for prompting both initial liking and securing the stable and long-term relationships that best satisfy belonging needs (Baumeister & Leary, 1995).

Based on these previous findings, we suggest that inclusion by in-group members creates perceptions of enhanced similarity with members of that group, an outcome we propose may be responsible for enhanced need satisfaction following inclusion by in-group members. Conversely, ostracism by in-group members may be interpreted as evidence of dissimilarity between the self and group, which may sharpen the negative experience of social exclusion. Specifically, to the extent that inclusion by in-group members makes the in-group more attractive to an individual, that person should indicate heightened perceived similarity with that in-group, resulting in higher levels of interpersonal satisfaction; conversely, ostracism by in-group members should lead individuals to perceive that group as less attractive, resulting in reduced perceptions of similarity with the in-group and potential interpersonal dissatisfaction (i.e., Morry, 2005). However, because out-groups are categorized as essentially dissimilar from the self, simple acts of inclusion and exclusion may lead to little change in perceived similarity to

out-group members, and thus may be unrelated to the experiences of ostracism and social inclusion.

To test our hypothesis, we manipulated the race of supposed confederates in the context of the Cyberball paradigm (Williams & Jarvis, 2006), thereby allowing us to independently manipulate whether participants experienced social inclusion or ostracism and whether that experience was initiated by members of one's own racial in-group or out-group. Following each of these experiences, we assessed participants' perceived similarity with the Cyberball "confederates" they played with as well as their basic needs satisfaction following the game. This design allowed us to replicate previous findings that inclusion and ostracism perpetrated by racial in-groups produce more intense interpersonal reactions (both positive and negative) as well as to test the hypothesis that the effects of in-group inclusion and ostracism on basic affiliation needs (dis)satisfaction are mediated by changes in perceived similarity with in-group members.

## Method

### Participants

The sample was composed of 67 (47 women;  $M_{\text{age}} = 18.97$  years,  $SD = 1.28$ ) Caucasian American undergraduate students at a large midwestern university in the United States who participated in exchange for partial course credit; non-Caucasian Americans did not participate in this study.<sup>1</sup> A 2 (inclusionary status: ostracism, inclusion)  $\times$  2 (agent race: same-race, other-race) between-subjects design was employed.

### Materials

**Basic Needs Scale.** To assess participants' reactions to social inclusion and ostracism, we administered the Basic Needs Scale, used extensively in prior research (Williams et al., 2000; Zadro, Williams, & Richardson, 2004). This questionnaire assesses four social needs known to fluctuate with experiences of social inclusion and ostracism: belonging, perceived

<sup>1</sup> No effects of participant sex were observed; therefore, this variable is not discussed further.



control, self-esteem, and meaningful existence (for a review, see Williams, 2007). The questionnaire included 16 items assessing perceived threats to these four needs (e.g., “I felt disconnected,” “I felt I had control over the situation,” “I felt liked,” “I felt meaningless”); we asked participants to provide their responses based on how they felt during the game (reflexive) rather than based on how they currently felt (reflective). Participants responded to each of the 16 questions using a 7-point Likert scale (1 = *not at all* to 7 = *extremely*), scored such that higher numbers indicated more fulfillment of the respective need (i.e., higher numbers indicated a greater experience of belonging, higher self-esteem, etc.).

**Perceived similarity.** Using the same 7-point Likert scale used to assess basic needs satisfaction, we also asked participants the extent to which they agreed with the statement “The other participants are similar to me.”

## Procedure

The procedure was adapted closely from that of Bernstein and colleagues (2010). Upon their arrival in a university laboratory, participants provided informed consent and were escorted to individual cubicles. Specifically, they were informed that they were participating in a study involving online communication and mental visualization, and that because the study was occurring in multiple departments on the university campus, they would be playing with other students located elsewhere on campus. In actuality, the other participants were computer-controlled agents programmed to either socially include or ostracize the participants on a between-subjects basis. To facilitate our experimental design, we told participants that they would be able to see a photograph of the two participants with whom they would be playing during the game itself, and that the two participants would be able to see a picture of them (thus allowing for our manipulation of agent race). Participants were then instructed to stand against a white wall while the experimenter took a digital photograph of them, which the experimenter then ostensibly uploaded to the computer network prior to the start of the game.

After having their picture taken and completing a series of filler questions supposedly measuring “mental visualization” capabilities, par-

ticipants were directed to play Cyberball, an Internet ball-tossing game ostensibly designed to help people hone their mental visualization skills. Importantly, our White participants saw either two White faces (i.e., in-group members) or two Black faces (i.e., out-group members) as fellow players (on a between-subjects basis, to manipulate racial in-group and out-group membership). In addition, participants were randomly assigned to either receive the ball roughly one third of the time (inclusion condition) or to receive the ball only twice at the beginning and then never again for the remaining throws (ostracism condition).

Participants then completed the similarity question and the Basic Needs Scale. Participants were debriefed and thanked for their participation. We used a funneled debriefing questionnaire to determine whether participants were suspicious about any aspect of our experimental procedures, including the Cyberball manipulation. Using an open-ended format, participants were asked, “Did you find anything odd or suspicious during the study?” and “Did you think anything in the study was fake?” If participants answered affirmatively to either question, they were asked to elaborate on their initial answer.

## Results

### Manipulation Check Effectiveness

To determine whether our Cyberball manipulation was believable to participants, we coded participants’ responses to our funneled debriefing questions for suspicion. Specifically, we created a “player suspicion” variable and noted any participants who indicated that they found the two Cyberball players generally or their race to be fake, odd, or suspicious. Anyone mentioning such a comment in either question was assigned a 1, and all others who did not mention anything were assigned a 0. This coding resulted in a total of six of 67 respondents having made comments regarding some aspect of the players being fake, odd, or suspicious. Importantly, only one participant’s suspicion was actually based on the race of the Cyberball players.

We conducted analyses to determine whether these differences varied across experimental conditions. We first conducted a chi-square

analysis, using Cyberball target race (Caucasian vs. African American) as the independent variable. This analysis yielded a nonsignificant difference,  $\chi^2(1) = 0.67, p = .41$ , indicating that suspicion about the players did not vary based on whether participants were playing with White or Black agents. We next conducted a similar chi-square analysis using Cyberball condition (social inclusion vs. ostracism) as the independent variable; again, this analysis revealed no difference between conditions,  $\chi^2(1) = 0.35, p = .56$ . To determine whether these factors (social experience and agent race) interacted, we conducted a binary logistic regression, dummy coding (1) inclusion + Black agents, ostracism + White agents, and ostracism + Black agents against the referent group of inclusion + White agents (0). The regression was nonsignificant,  $\chi^2(3) = 2.82, p = .42$ , for each of the three dummy-coded variables entered into the logistic regression.

In a final analysis, we dummy coded both social experience (0 = exclusion, 1 = inclusion) and agent race (0 = Black, 1 = White) and entered them simultaneously as factors in a binary logistic regression, using our “player suspicion” factor as the dependent variable. This analysis allowed us to examine the effect of each of these factors while also accounting for the variability due to the other factor. This analysis also yielded a nonsignificant regression,  $\chi^2(2) = 0.98, p = .61$ .

All evidence suggested that suspicion about the Cyberball players, including their race, did not differ between conditions of our experiment. Furthermore, this identical experimental procedure has been used in previous work and similarly low levels of suspicion were documented (see Bernstein et al., 2010). As such, the current procedure possessed an adequate level of experimental realism.

### Basic Needs

The individual needs were highly related ( $\alpha = .94$ ); as such, we formed a composite basic needs score for each participant (higher values represent more satisfied basic needs; see McConnell, Brown, Shoda, Stayton, & Martin, 2011, for similar procedures). We then subjected these scores to a 2 (inclusionary status: inclusion vs. ostracism)  $\times$  2 (agent race: same-race vs. other-race) between-subjects analysis

of variance (ANOVA).<sup>2</sup> A main effect of inclusionary status emerged,  $F(1, 63) = 105.73, p < .001, \eta_p^2 = .63$ , such that included participants ( $M = 4.44, SE = 0.14$ ) reported substantially more fulfilled basic needs than did ostracized participants ( $M = 2.50, SE = 0.13$ ). Importantly, this main effect was qualified by a significant Inclusionary Status  $\times$  Agent Race interaction,  $F(1, 63) = 5.41, p = .023, \eta_p^2 = .079$  (see Figure 1). Simple effects analyses revealed that participants ostracized by same-race agents ( $M = 2.22, SD = 0.58$ ) had less fulfilled basic needs than did those ostracized by other-race agents ( $M = 2.78, SD = 0.88$ ),  $t(35) = 2.20, p = .032, d = 0.75$ . Although participants included by same-race agents were directionally more satisfied ( $M = 4.61, SD = 0.83$ ) than were those included by other-race agents ( $M = 4.28, SD = .73$ ), this effect did not reach statistical significance,  $t(28) = 1.15, p > .20, d = 0.42$ . This significant interaction between inclusionary status and target race, however, is consistent with previous findings indicating that although ostracism always feels worse than social inclusion, such effects can be qualified by the in-group–out-group relationship of the rejecter and rejected (Bernstein et al., 2010).

### Mediation Analysis

We then investigated whether the stronger responses to in-group inclusion and ostracism were mediated by changes in the perceived similarity of the self to in-group members.<sup>3</sup> We predicted that participants included by same-race participants should see the self as more similar to the agents, relative to participants experiencing ostracism. We further predicted that increased perceived similarity should be positively correlated with increased fulfillment of basic affiliation needs. Finally, we predicted that the relationship between inclusionary status and basic needs would be mediated by the

<sup>2</sup> Treating each of the four subscales as separate dependent variables in four separate ANOVAs had no effect on the interaction pattern ( $ps < .07$ ).

<sup>3</sup> Because perceived similarity and belongingness are potentially psychologically similar constructs, we conducted a correlational analysis to determine how strongly they were related to one another,  $r(67) = .49, p < .01$ . Because these two concepts were only moderately correlated, perceived similarity was an adequate variable to include as a psychological mediator.

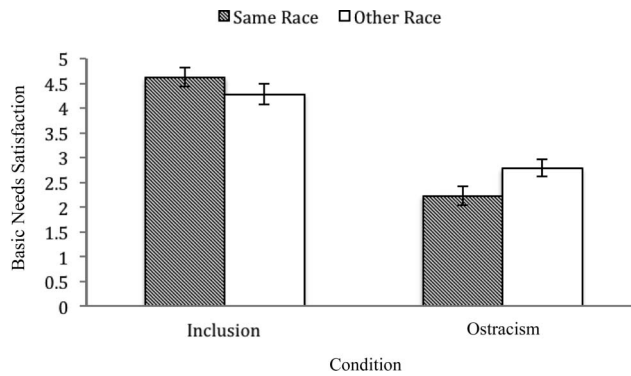


Figure 1. The effect of inclusion and ostracism on basic needs satisfaction as a function of the in-group–out-group status of the agents.

aforementioned increase in perceived similarity. However, we believe this relationship should only be true for participants playing with same-race agents. We hypothesized that de facto dissimilarity with racial out-groups would be unaffected by the relatively minor inclusionary and exclusionary behaviors used in our task, and as such, we did not expect similarity to mediate the relationship between inclusionary status and basic needs satisfaction.

We therefore conducted a moderated mediation analysis (i.e., a conditional indirect effect; see Preacher, Rucker, & Hayes, 2007) to examine our specific hypothesis, and we used the Preacher et al. (2007) existing SPSS macro to conduct this analysis. Before testing the moderated mediation, we began by dummy coding the independent variable of inclusionary status (1 = inclusion, 0 = ostracism) and entered it as a predictor of the composite basic needs score. We also included our similarity measure as our mediator and target race as the moderating variable (0 = White, 1 = Black). This analysis produced a significant overall model,  $F(3, 63) = 5.82, p = .014, R^2 = .22$ , which included our previously described direct effect on the basic needs, such that inclusion increased basic needs satisfaction relative to ostracism ( $B = 1.94, p < .001$ ). We next examined whether the direct effect was mediated by perceptions of similarity. To test this hypothesis, we employed the bootstrapping method developed by Preacher and Hayes (2004, 2008). This method requires that a relationship between an independent variable and a dependent variable exist to be mediated, and that the “indirect effect be

statistically significant in the direction predicted by the mediation hypothesis” (Preacher & Hayes, 2004, p. 719). Having already established these requirements, we next performed the bootstrapping procedures using the macros provided by Preacher and Hayes (2008) to determine whether the direct effect of inclusionary status on basic needs was mediated by perceptions of similarity. This analysis yielded a 95% bias-corrected confidence interval (based on 5,000 bootstrap samples) for the indirect effect that did not include zero (95% CI [0.066, 0.70]). This analysis indicated that perceptions of perceived similarity mediated the direct effect. Greater social inclusion led to stronger perceptions of similarity ( $B = 0.92, p = .002$ ), and as similarity increased, so too did participants’ basic needs satisfaction ( $B = 0.28, p < .001$ ). The direct effect, however, was mediated when including similarity ( $B = 1.68, p < .001$ ).

However, we wished to examine whether this was true both when playing with same-race and other-race agents (i.e., did target race moderate the mediated effect?). We conducted a moderated mediation analysis using Model 7 with the W moderator being target race (0 = White, 1 = Black). This yielded two bootstrapping models, each with 5,000 iterations. When participants played with White agents, similarity did indeed mediate the effect of inclusion and ostracism on basic needs satisfaction (95% bias-corrected confidence interval did not include zero; 95% CI [0.079, 0.91]). However, similarity did not mediate the effect when participants played with Black agents (95% bias-corrected confidence interval did include zero; 95% CI [−0.

054, 0.39]). Thus, our moderated mediation hypothesis was supported.

### Discussion

Consistent with past research, we found that the social pain stemming from ostracism is more severe when administered by fellow in-group members (e.g., [Bernstein et al., 2010](#)). Conversely, social inclusion is experienced as relatively more positive when administered by in-group relative to out-group members. Importantly, we found that these enhanced effects of social inclusion and ostracism observed in the in-group condition are mediated by perceived similarity with those agents of inclusion or ostracism. Not only do in-group members seem more similar to the self when they include participants (and more dissimilar when they ostracize participants), but changes in the perception of similarity with agents mediate the relationship between inclusionary status and the fulfillment of basic affiliation needs. Critically, this relationship was only true when participants played the game with racial in-group members; perceptions of similarity did not change, regardless of the inclusionary experience, in the racial out-group condition (i.e., moderated mediation).

Several theoretical positions are congruent with our findings. First, out-group members are naturally constrained with respect to how similar they may be perceived to in-group members. Among racial out-groups, for whom the signals of group differences (i.e., skin tone) are salient (e.g., [Ito & Urland, 2003](#)), it may be difficult to ever completely overcome that difference so long as “race” is the primary category that defines the targets. This explanation is further supported by evidence showing that out-groups are perceived as being more homogenous as compared with in-groups ([Park, Ryan, & Judd, 1992](#)). It may be that when individuals think of in-group members, they are better able to imagine a greater degree of variation in how similar they are to themselves; receiving positive feedback (such as acceptance) may enhance these perceptions of similarity, because positive interpersonal interactions can increase perceived similarity ([Morry, 2005](#)). At the same time, receiving negative feedback may result in perceiving the agent as less similar to themselves (perhaps the group as well) as a means of dis-

tancing oneself from the aggressor and perhaps also protecting generally positive perceptions of the in-group (e.g., [Doosje, Ellemers, & Spears, 1995](#)). Alternatively, ostracism from the in-group may threaten one’s perceived standing within the in-group ([Williams, 2007](#)), causing the observed dissimilarity and intensified negativity.

Although past theory and research (e.g., [Williams, 2007](#)) proposed that early stage reactions to ostracism are essentially reflexive and insensitive to situational factors, the current results and other recent work ([Bernstein et al., 2010](#)) suggest that responses to ostracism are not simply “all or nothing,” and instead can be more nuanced and contextualized. Furthermore, the current results also mesh nicely with social pain theory ([MacDonald & Leary, 2005](#)), which proposes that social and psychical pain are phenomenologically and physiologically linked experiences (e.g., [Eisenberger, Lieberman, & Williams, 2003](#)). To the extent that reactions to ostracism are similar to responses to physical pain, it is sensible that the most serious social injuries (e.g., being ostracized by in-groups) should result in the most acute experiences of social pain.

Nonetheless, a number of open questions still remain. First, given that the current work used race as the operational definition of in-group and out-group, it may appear unclear whether the current effects are due to an in-group effect or due to status or power differences among the in-group and out-group. Although a sensible alternative explanation, this hypothesis is not easily congruent with previous findings. Indeed, [Bernstein and colleagues \(2010\)](#) found that both Black and White participants showed equivalent effects when socially included or ostracized by racial in-group versus out-group members. Clearly, this pattern of data does not align easily with a status or power interpretation of race. Second, in the current work, we did not test whether similar results would be observed for other in-group–out-group distinctions. Here, too, we can point to past research to address this issue. Given that the differential responses to ostracism and social inclusion by in-group and out-group members only appear to occur for highly essentialized groups (e.g., race; see [Bernstein et al., 2010](#)), but not for more arbitrary groups or groups with weak affiliations (e.g., PC vs. Mac users; [Smith & Williams,](#)

2004), this indicates that the current mediational pattern for in-groups would likely hold only for subjectively relevant and highly essentialized in-groups.

In conclusion, consistent with past research, we found that reactions to social inclusion and ostracism are exacerbated when perpetrated by fellow racial in-group members (compared with racial out-group members). Importantly, these exacerbated reactions stemming from in-group inclusion and ostracism are explained by fluctuations in perceived self-similarity with in-group members. Future research would benefit by understanding additional mechanisms potentially responsible for differentiated reactions to inclusion and ostracism by in-group members.

### References

- Acitelli, L. K., Kenny, D. A., & Weiner, D. (2001). The importance of similarity and understanding of partners' marital ideals to relationship satisfaction. *Personal Relationships*, 8, 167–185. doi:10.1111/j.1475-6811.2001.tb00034.x
- Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, 117, 497–529. doi:10.1037/0033-2909.117.3.497
- Bernstein, M. J., Sacco, D. F., Young, S. G., Hugenberg, K., & Cook, E. (2010). Being “in” with the in-crowd: The effects of social exclusion and inclusion are enhanced by the perceived essentialism of the ingroup and outgroup. *Personality and Social Psychology Bulletin*, 36, 999–1009. doi:10.1177/0146167210376059
- Correll, J., & Park, B. (2005). A model of the ingroup as a social resource. *Personality and Social Psychology Review*, 9, 341–359. doi:10.1207/s15327957pspr0904\_4
- Doosje, B., Ellemers, N., & Spears, R. (1995). Perceived intragroup variability as a function of group status and identification. *Journal of Experimental Social Psychology*, 31, 410–436. doi:10.1006/jesp.1995.1018
- Eisenberger, N. I., Lieberman, M. D., & Williams, K. D. (2003, October 10). Does rejection hurt? An fMRI study of social exclusion. *Science*, 302, 290–292. doi:10.1126/science.1089134
- Feingold, A. (1988). Matching for attractiveness in romantic partners and same-sex friends: A meta-analysis and theoretical critique. *Psychological Bulletin*, 104, 226–235. doi:10.1037/0033-2909.104.2.226
- Ferguson, C. K., & Kelley, H. H. (1964). Significant factors in overevaluation of own-group's product. *Journal of Abnormal Psychology*, 69, 223–228. doi:10.1037/h0046572
- Gonsalkorale, K., & Williams, D. (2007). The KKK won't let me play: Ostracism even by a despised outgroup hurts. *European Journal of Social Psychology*, 37, 1176–1186. doi:10.1002/ejsp.392
- Gonzaga, G. C., Campos, B., & Bradbury, T. (2007). Similarity, convergence, and relationship satisfaction in dating and married couples. *Journal of Personality and Social Psychology*, 93, 34–48. doi:10.1037/0022-3514.93.1.34
- Goodwin, S. A., Williams, K. D., & Carter-Sowell, A. R. (2010). The psychological sting of stigma: The costs of attributing ostracism to racism. *Journal of Experimental Social Psychology*, 46, 612–618. doi:10.1016/j.jesp.2010.02.002
- Hein, G., Silani, G., Preuschhoff, K., Batson, C. D., & Singer, T. (2010). Neural responses to ingroup and outgroup members' suffering predict individual differences in costly helping. *Neuron*, 68, 149–160. doi:10.1016/j.neuron.2010.09.003
- Ito, T. A., & Urland, R. (2003). Race and gender on the brain: Electrocortical measures of attention to the race and gender of multiply categorizable individuals. *Journal of Personality and Social Psychology*, 85, 616–626. doi:10.1037/0022-3514.85.4.616
- Jensen, A. R. (1978). Genetic and behavioral effects of nonrandom mating. In R. T. Osborne, C. E. Noble, & N. J. Weyl (Eds.), *Human variation: Biopsychology of age, race, and sex* (pp. 51–105). New York, NY: Academic Press.
- Leary, M. R., Haupt, A. L., Strausser, K. S., & Chokel, J. T. (1998). Calibrating the sociometer: The relationship between interpersonal appraisals and state self-esteem. *Journal of Personality and Social Psychology*, 74, 1290–1299. doi:10.1037/0022-3514.74.5.1290
- MacDonald, G., & Leary, M. R. (2005). Why does social exclusion hurt? The relationship between social and physical pain. *Psychological Bulletin*, 131, 202–223. doi:10.1037/0033-2909.131.2.202
- McConnell, A. R., Brown, C. M., Shoda, T. M., Stayton, L. E., & Martin, C. E. (2011). Friends with benefits: On the positive consequences of pet ownership. *Journal of Personality and Social Psychology*, 101, 1239–1252. doi:10.1037/a0024506
- Morry, M. M. (2005). Relationship satisfaction as a predictor of similarity ratings: A test of the attraction-similarity hypothesis. *Journal of Social and Personal Relationships*, 22, 561–584. doi:10.1177/0265407505054524
- Park, B., Ryan, C. S., & Judd, C. M. (1992). Role of meaningful subgroups in explaining differences in perceived variability for in-groups and out-groups. *Journal of Personality and Social Psychology*, 63, 553–567. doi:10.1037/0022-3514.63.4.553

- Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, *36*, 717–731. doi:10.3758/BF03206553
- Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*, 879–891. doi:10.3758/BRM.40.3.879
- Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research*, *42*, 185–227. doi:10.1080/00273170701341316
- Prentice, D. A., & Miller, D. T. (2007). Psychological essentialism of human categories. *Current Directions in Psychological Science*, *16*, 202–206. doi:10.1111/j.1467-8721.2007.00504.x
- Russell, R. J., & Wells, P. A. (1991). Personality similarity and quality of marriage. *Personality and Individual Differences*, *12*, 407–412. doi:10.1016/0191-8869(91)90057-I
- Smith, A., & Williams, K. D. (2004). R U there? Effects of ostracism by cell phone messages. *Group Dynamics: Theory, Research, and Practice*, *8*, 291–301. doi:10.1037/1089-2699.8.4.291
- Tajfel, H. (1982). Social psychology of intergroup relations. *Annual Review of Psychology*, *33*, 1–39. doi:10.1146/annurev.ps.33.020182.000245
- Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, *1*, 149–178. doi:10.1002/ejsp.2420010202
- Williams, K. D. (2007). Ostracism. *Annual Review of Psychology*, *58*, 425–452. doi:10.1146/annurev.psych.58.110405.085641
- Williams, K. D., Cheung, C. K. T., & Choi, W. (2000). Cyberostracism: Effects of being ignored over the Internet. *Journal of Personality and Social Psychology*, *79*, 748–762. doi:10.1037/0022-3514.79.5.748
- Williams, K. D., & Jarvis, B. (2006). Cyberball: A program for use in research on ostracism and interpersonal acceptance. *Behavior Research Methods*, *38*, 174–180. doi:10.3758/BF03192765
- Williams, K. D., & Zadro, L. (2005). Ostracism: An indiscriminate early detection system. In K. D. Williams, J. P. Forgas, & W. von Hippel (Eds.), *The social outcast: Ostracism, social exclusion, rejection, and bullying* (pp. 19–34). New York, NY: Psychology Press.
- Zadro, L., Boland, C., & Richardson, R. (2006). How long does it last? The persistence of the effects of ostracism in the socially anxious. *Journal of Experimental Social Psychology*, *42*, 692–697. doi:10.1016/j.jesp.2005.10.007
- Zadro, L., Williams, K. D., & Richardson, R. (2004). How low can you go? Ostracism by a computer is sufficient to lower self-reported levels of belonging, control, self-esteem, and meaningful existence. *Journal of Experimental Social Psychology*, *40*, 560–567. doi:10.1016/j.jesp.2003.11.006

Received June 22, 2013

Revision received August 27, 2013

Accepted August 27, 2013 ■

### E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at <http://notify.apa.org/> and you will be notified by e-mail when issues of interest to you become available!

## Single-Case Designs in Group Work: Past Applications, Future Directions

Mark J. Macgowan and Stephen E. Wong

Robert Stempel College of Public Health & Social Work, Florida International University

This paper examined how group work researchers and practitioners have used single-case designs (SCDs) to evaluate interventions for improving group processes and outcomes. Fifty-one group work studies published from 1972 to 2012 using SCDs and addressing a wide range of problems were identified. The 8 types of SCDs applied in these studies are discussed in relation to their methodology, relative strengths and limitations, and applications in group work research and practice evaluation. Selected studies, which varied in quality and rigor, are described to illustrate design features and utility for group work. The review reveals the challenges and opportunities in using SCDs, and it provides recommendations for designs best suited to answer particular questions.

**Keywords:** group work, single-case designs, single-subject designs, single-system designs, practice evaluation

Single-case designs (SCDs), also known as *single-subject*, *single-participant*, *N-of-1*, and *single-system evaluation designs*, can evaluate effects of an intervention with only one subject or group. In most SCDs, repeated measures are taken on a single subject or group under no-treatment (baseline) and treatment conditions, and the subject or group serves as its own control. Controlled SCDs demonstrate experimental control by replicating treatment effects within each single-case study through procedures such as repeatedly introducing and removing the intervention; successively introducing the intervention across different targets, settings, or subjects; randomly alternating between various conditions; or incrementally changing the level of the intervention. Results of SCDs are typically graphed for visual analysis, but there are also

numerous tests for determining statistically significant change.

SCDs have been structured to examine special features of the group modality. The design may focus on functioning of individuals within the group, a subgroup of individuals, or the whole group. SCDs can also include variables that reflect the group experience, such as member satisfaction, nature of interactions, and leadership. There are many potential focal points such as intervention theory, individual group members, group structures, group processes and leadership, all of which relate to group work outcomes (Burlingame, Strauss, & Joyce, 2013).

### Benefits of Single-Case Designs for Group Work

SCDs are considered a useful research design for clinical practice across professions including psychology, social work, special education, and counselor education (Bloom, Fischer, & Orme, 2009; Heppner, Kivlighan, & Wampold, 2008; Horner et al., 2005; Kazdin, 2003; Lundervold & Belwood, 2000; Vonk, Tripodi, & Epstein, 2007). They can be a valid source of data in determining empirically supported interventions (Chambless et al., 1998). In group work it has been recommended as a valuable

---

Mark J. Macgowan and Stephen E. Wong, Robert Stempel College of Public Health & Social Work, Florida International University.

Both authors contributed equally to the writing of this paper and are listed in alphabetical order.

Correspondence concerning this article should be addressed to Mark J. Macgowan or Stephen E. Wong, School of Social Work, Robert Stempel College of Public Health and Social Work, Florida International University, 11200 SW 8th Street, Miami, FL 33199. E-mail: [macgowan@fiu.edu](mailto:macgowan@fiu.edu); [wongse@fiu.edu](mailto:wongse@fiu.edu)

evaluation method (Garrett, 2005; Toseland & Rivas, 2012; Zastrow, 2006).

There are a number of important advantages of SCDs for research and practice. SCDs offer a means of obtaining valuable pilot data before larger between-groups designs are utilized (Petry, Martin, & Finocche, 2001; Petry, Weinstock, & Alessi, 2011). Between-groups designs require a control group and a sufficient number of subjects to satisfy the power requirements of the statistical test to be used for data analysis. However, SCDs can be conducted with a single person or one group containing as many members who attend. SCDs are also the most practical technology available for evaluating one's own practice, which is required by the American Group Psychotherapy Association (AGPA, 2007), Association for Specialists in Group Work (ASGW, 2000, 2007), and the Association for the Advancement of Social Work with Groups (AASWG, 2006). Another benefit of SCDs is their flexibility, as they can accommodate a trial-and-error process allowing clinicians to try out different group interventions in the search for an effective one (Newman & Wong, 2004). This also makes SCDs particularly well suited for practitioners using evidence-based group treatments with diverse populations. Group workers using SCDs can determine whether an evidence-supported intervention or technique, validated with other populations, is effective with the particular cultural groups they are serving (Macgowan, 2008; Macgowan & Hanbidge, 2014).

### Previous Reviews

Compared with between-groups designs, SCDs seldom appear in published group work research. Reviews of the history of group work research since the 1950s made no mention of SCDs (Barlow, Burlingame, & Fuhriman, 2000; Barlow, Fuhriman, & Burlingame, 2004; Feldman, 1987; Frank, 1975; Horne & Rosenthal, 1997; McGrath, 1997). In a review of 54 studies in a decade of group work research up to 1994, only two SCDs were identified (Tolman & Molidor, 1994). As the authors noted, the lack of studies might have been a reflection of "publication biases against such studies, rather than researcher rejection of such designs" (Tolman & Molidor, 1994, p. 152).

Regardless of the reasons, the focus in group research has been on between-groups designs, with little attention on SCD technology. With respect to clinical practice, relatively few practitioners use SCDs, but instead rely on informal methods to determine progress (Foster, Watson, & Young, 2002; Galassi & Gersh, 1993; Ventimiglia, Marschke, Carmichael, & Loew, 2000). In group work, the findings are similar. A survey of the use of SCDs by 54 school practitioners found that none used SCDs in their groups but rather relied on subjective methods (Garrett, 2005, p. 251).

There have been a number of reviews of SCDs in group work (Fike, 1980; Johnson, Beckerman, & Auerbach, 2001; Patterson & Basham, 2002; Robison, Morran, & Hulse-Killacky, 1989; Rose, 1977), a few of which can be summarized. In his book on behavioral group work, Rose (1977, pp. 66–72) described seven applications of SCDs, illustrating five simple baseline-treatment designs (AB), one reversal design (ABAB), and one multiple-baseline design (MBD). These examples illustrated how to use SCDs to assess change in variables such as member-to-member interactions, member praise statements, leader-to-member interactions, leader suggestion giving, positive affective responses, suggestion-opinion giving responses, information-giving responses, and subjective appraisal of intervention procedures by group members.

Fike (1980) noted how SCDs could be used flexibly to capture the nature of change in groups. He described the possible relationships between multiple formative/process variables and outcome variables in groups. He argued that designs simply recording individual outcomes, "do not give attention to all of the relevant group information" (Fike, 1980, p. 42). According to Fike, SCDs should also include other group variables such as member satisfaction with the group, quality of interactions, and group leadership.

Robison and colleagues (Robison et al., 1989) described two SCDs (ABAB and MBD) and statistical procedures (time-series analyses and nonparametric tests) that group workers could use to evaluate the effectiveness of their interventions. They cautioned that ABAB designs might not be able to determine causality with some dependent variables, such as cohesion, as the developmental processes may occur irrespec-



tive of an intervention and might not be reversible through withdrawal of treatment. In such cases, a MBD across clients or groups could be used to demonstrate causality. To improve the internal validity of findings from a multiple-baseline across groups design, the authors recommended random assignment of clients into the different groups.

### Purpose and Scope of Review

This paper advances previous reviews by including more recent studies; examining a broader range of SCD designs, their individual methodologies, and relative strengths and weaknesses; and making recommendations for practice and research with groups. This review is restricted to SCDs involving facilitated small groups whose purpose is to improve the psychosocial or socioemotional functioning of its members. The term *small group* “implies the ability of members to identify themselves as members, to engage in interaction, and to exchange thoughts and feelings among themselves through verbal, nonverbal, and written communication processes. Members can meet face-to-face, by telephone or video, or through computer networks” (Toseland & Rivas, 2012, p. 11).

We used a number of methods to locate and identify SCDs in the literature. We restricted the search to scholarly journals of all years, up to 2012. We searched a number of databases including social service abstracts, PsycINFO, Sociofile, Medline, and Social Work Abstracts. The reference lists of other reviews (Garrett, 2005; Johnson et al., 2001; Maughan, Christiansen, Jenson, Olympia, & Clark, 2005; Smith, 2012) were also examined, in addition to those in recent SCD studies. The following terms related to group work and SCDs were used to locate the literature, searching in the title, keywords, and abstract: single subject, single system, single case, multiple baseline, withdrawal, reversal, alternating treatments design, AB, ABA, BAB, ABAB, appearing with any of the following terms: group work, group counseling, group treatment, group intervention, group therapy, group prevention, group approach, group psychotherapy, group training, and group instruction. After an iterative process to select studies meeting the above criteria, 51 studies were found.

In Table 1 and the following section we review SCDs used in extant group work research, beginning with the simplest and progressing to the more complex designs. We briefly describe the logic behind each design, what the design can accomplish, and how it was properly or less-than optimally applied in previous studies. We also note when these designs were used to investigate particular questions about group processes or outcomes. Discussion and recommendations for research and practice follow.

### Review of Single-Case Designs in Group Work Studies

The first SCD in group work appeared in 1972 with an additional five published in that decade (see Table 1). There were 15 published in the 1980s, 11 in the 1990s, and 19 since the year 2000, representing a considerable increase from the previous decade. The discipline with the most publications is Psychology followed by Social Work. Most of the studies have originated in the United States but studies were also conducted in Argentina, New Zealand, and the United Kingdom.

#### B Design (Treatment Only)

Four of the publications in Table 1 used a B design, which first appeared in the 1980s. It is the simplest SCD and it is uncontrolled, consisting of merely taking repeated measurements during treatment. When outcome measures are collected in this manner, various results are possible: (a) measures might change in the desired direction, indicating a clinical improvement; (b) measures might not change significantly, indicating a lack of improvement despite treatment; (c) measures might change in the undesired direction, indicating clinical deterioration; or (d) some intermediate result that falls between a and b or between b and c. These results are important for deciding whether to continue treatment and for making future treatment recommendations, making the B design valuable for clinicians. However, although the B design is easy to do and it is superior to unsystematic or no collection of outcome data, it has serious limitations. If the client does improve during treatment, numerous alternative explanations such as history, maturation, testing (Campbell & Stanley, 1966; Cook & Campbell,

Table 1  
Single-Case Designs Previously Applied in Group Work Research

Design	Brief Description	Advantages	Disadvantages	Considerations for Group Work	Studies Using this Design
B	Monitor outcome measures during treatment. Graphic display can show improvement, no change, or worsening associated with intervention.	Simplest & easiest SCD to conduct. Does not require a no-treatment baseline phase.	Unable to draw causal inferences. Susceptible to confounds from history, maturation, testing, natural healing, & extraneous events.	Previous researchers have used this design to monitor if formative group processes are occurring as expected.	(Cetingok & Hirayama, 1983*; Fontao & Mergenthaler, 2008*; O'Brien et al., 1999*; Patterson & Basham, 2002)
AB & related designs	After collecting stable baseline data or data moving in a counter-therapeutic direction, apply the intervention for a period at least equal to the length of baseline.	Relatively easy & practical to conduct.	Sometimes not possible to delay treatment to collect baseline data. Substantial changes evident in a single AB design can only suggest a treatment effect. Should not make causal inferences from this design because it is susceptible to confounds from extraneous events.	Must consider the possibility that emergent changes coinciding with treatment are due to developmental or internal group processes.	(Hall, 2006; Johnson, et al., 2001*; Ledgerwood, et al., 2008 [counterbalanced AB]*; Nelson & Kelley, 2001; Rubin, 1991*; Stevens, 2006; Toseland, Krebs, & Vahsen, 1978; Wade, Ortiz, & Gorman, 2007 [interrupted time-series design])
ABC, ABCD AB <sup>1</sup> B <sup>2</sup> , AB <sup>1</sup> B <sup>2</sup> B <sup>3</sup> , & related designs	When initial treatment (B or B <sup>1</sup> ) proves unsatisfactory, apply a second treatment (C) or the same treatment at a higher intensity (B <sup>2</sup> ). If B <sup>2</sup> proves unsatisfactory, apply a third treatment (D) or the same treatment at an even higher intensity (B <sup>3</sup> ).	Accommodates a flexible trial & error process. A series of interventions can be applied until a satisfactory outcome is attained.	Like an AB design, is susceptible to confounds from extraneous events with the seemingly effective treatment. Additional threats to this design include carryover effects from prior treatments.	Developmental & internal group processes may also confound the internal validity of these designs.	(Fry, et al., 2009 [ABC]*; Phaneuf & McIntyre, 2011 [ABCD]*)

(table continues)

Table 1 (*continued*)

Design	Brief Description	Advantages	Disadvantages	Considerations for Group Work	Studies Using this Design
ABA, BAB, ABAB, ABCBC, AB <sup>1</sup> B <sup>2</sup> B <sup>1</sup> B <sup>2</sup> , & other reversal designs	Remove an apparently effective treatment or return to the condition that preceded the apparently effective treatment to reverse observed improvements, & then reintroduce the treatment to reproduce and replicate its effects.	Experimental & Controlled Designs Replication of treatment effects provides a strong demonstration of experimental control over the dependent variable (good internal validity). Design can be conducted with only one participant or one group.	Ethical concerns about intentionally reversing gains associated with a treatment. Not applicable with some interventions (e.g., psychoeducational) because their effects are not reversible.	Certain group processes (e.g., cohesion, role differentiation) are unlikely to be reversible making the design unworkable.	(Hauserman, et al., 1972 [ABAB]); Helmus, et al., 2003 [ABA]; Kirby, et al., 2008; Linsk, et al., 1975 [ABAB] <sup>*</sup> ; Marr & Fairchild, 1993 [ABAA]; Martinez & Wong, 2009 [ABAB] <sup>*</sup> ; Mouton & Stanley, 1996 [ABA]; Petry, et al., 2001 [AB <sup>1</sup> B <sup>2</sup> CA]; Vaccaro, 1990 [ABAB]; Weber, 1980 [ABAB] <sup>*</sup> )
Concurrent multiple-baseline design	Simultaneously collect data on two or more equivalent baselines. After obtaining stable pre-intervention data, apply the intervention to the first baseline. After observing clear gains on the first baseline, apply the intervention to the second baseline. Repeat this process with third & subsequent baselines.	Design does not require a reversal & therefore is not complicated by possible ethical concerns & irreversibility. As the treatment is replicated across different behaviors, members, or groups, evidence for the generalizability (external validity) of the intervention is obtained.	Design requires two or more independent baselines (i.e., similar behaviors, settings, participants, or groups). Intervention is delayed for second & subsequent baselines.	MBD across participants within the same group is vulnerable to carry-over across members due to modeling effects & changes in group interaction patterns. Generalization across participants or groups cannot be distinguished from loss of experimental control (i.e., extraneous effects).	(Bates, 1980; Clark, Cunningham, & Cunningham, 1989; Foxx, McMorrow, Bittle, & Ness, 1986; Foxx, et al., 1983; Gronna, Serna, Kennedy, & Prater, 1999; Hall, et al., 2000 <sup>*</sup> ; Hall, Schlesinger, & Dineen, 1997; Hansen, et al., 1985 <sup>*</sup> ; Hansen, St. Lawrence, & Christoff, 1989;

Table 1 (continued)

Design	Brief Description	Advantages	Disadvantages	Considerations for Group Work	Studies Using this Design
Noncurrent multiple-baseline design	The NCMBD resembles a succession of AB designs. In its strictest form the length of the different baselines vary & are randomly assigned to a series of clients or groups.	Design does not require running multiple baselines simultaneously. The NCMBD replicates treatment effects across participants or groups & therefore demonstrates generality with each successive baseline.	Requires multiple participants or groups.	Because they are started at different points in time, there is less likelihood of carryover effects across participants & groups.	Kelly, Laughlin, Claiborne, & Patterson, 1979; Kelly, Wildman, Urey, & Thurman, 1980; Lumpkin, Silverman, Weems, Markham, & Kurtines, 2002 [partial nonconcurrent MBD]; Phaneuf & McIntyre, 2007; Roessler & Lewis, 1984; Serna, Schumaker, Sherman, & Sheldon, 1991; Tracey, et al., 1974; Tymchuk, Andron, & Rahbar, 1988) (Connors, et al., 1984*; Edleson, et al., 1985*; Kettlewell, et al., 1992*; Lees & Roman, 2008)

(table continues)

Table 1 (continued)

Design	Brief Description	Advantages	Disadvantages	Considerations for Group Work	Studies Using this Design
Alternating treatments design	Baseline sessions & treatment sessions are rapidly alternated according to a random schedule.	Efficient design that allows the comparison of multiple treatments within a short period of time. Can show treatment effects even when baseline data is trending in the desired direction. Brief presentation of treatments may minimize irreversibility & multiple treatment interference.	Difficult to use with complex therapy protocols that involve a cumulative building process. Given rapidly alternating treatments, there is the risk of interaction effects with typical group development events.	Can be used to test effects of interventions on outcomes & for comparing the effects of interventions on within-group conditions (e.g., engagement).	(Blake, et al., 1990 <sup>*</sup> ; Gajar, et al., 1984; Kastner & May, 2009 <sup>*</sup> ; Wong, et al., 1988 <sup>*</sup> )
Mixed or combined designs	Combining SCDs or embedding one SCD within another can increase design manipulations to provide more demonstrations of experimental control & answer complex questions.	Can permit more sophisticated inquiry & refinement of interventions (e.g., after demonstrating an effective therapy assessing strategies to promote generalization).	Mixed & combined designs require higher proficiency in the use of SCD methodology.	Complex designs could be structured to produce sequential changes in group processes & group outcomes to test the relationships between those variables.	(Kohler & Fowler, 1985 <sup>*</sup> ; Wong, et al., 1996 <sup>*</sup> )

\* Article is further reviewed in the text of this paper.

1979), and natural healing (Wong, 2010) prevent one from concluding that treatment caused those improvements.

Group process measures can also be plotted during a B design, revealing fluctuations in variables such as verbal interaction patterns and group cohesion, which might mediate outcomes for group members. A few group researchers have monitored group process during treatment in this manner (Fontao & Mergenthaler, 2008; O'Brien, Korchynsky, Fabrizio, McGrath, & Swank, 1999; Patterson & Basham, 2002). For example, O'Brien and colleagues (1999) measured positive process, satisfaction, and negative process during stress management groups, and reported the findings to the group. In addition, the intervention included increasing positive interactions (e.g., active helping) during the latter sessions of the group program. The researchers found that ratings of positive process and satisfaction increased significantly across sessions. These investigators further reported that satisfaction ratings were inversely correlated to cardiovascular reactivity to stress, while negative process ratings were positively correlated to this physiological measure. The authors speculated that the planned emphasis on positive interaction may have improved ratings over time. More recently, Fontao and Mergenthaler (2008) examined the verbal content of female patients at four measurement points in a psychodynamic group for eating disorders. As predicted by the researchers, group members' language patterns ("emotion-abstraction patterns") significantly related to particular therapeutic factors, which also suggested the presence of a positive therapeutic cycle. The researchers noted that such language patterns may be used to identify the occurrence of therapeutically important points in the process of group work.

Although process measures can serve as proximal indicators of change and are important for assessing whether desired therapeutic processes unfold as planned, the ultimate test of group therapy is showing improvement in client outcomes. Cetingok and Hirayama (1983) used a B design consisting of four assessments conducted at the first, third, fifth, and seventh weeks of their program to assess whether participation in groups was associated with gains in health care knowledge and mental health status for elderly persons with diabetes, heart disease, and hypertension. The investigators found no

significant gains in any of their dependent measures. Cetingok and Hirayama explained their unexpected findings by noting that their elderly subjects probably entered the program well informed about their health care needs and apparently became suspicious that the mental health assessment data might be used to expel them from their rent-subsidized, high-rise apartments.

### **AB Design (Baseline; Treatment)**

The AB design is a substantial improvement over the B design and unlike the latter could be considered an actual "design" in that it provides a minimal degree of experimental control. Control comes from the baseline data (during which no treatment is administered) to which treatment data is compared. If the baseline data are stable or moving in a countertherapeutic direction and this pattern changes shortly after the introduction of the intervention, these results can give evidence that the intervention might be effective. Unlike the B, the AB design eliminates maturation (i.e., processes associated with the mere passage of time) and testing (i.e., measurement reactivity) as alternative explanations of observed change (Campbell & Stanley, 1966). Unfortunately, the AB design is still susceptible to "history" or extraneous events that happen to coincide with treatment (Campbell & Stanley, 1966; Hersen & Barlow, 1976), such as the arrival or departure of group members. Eight studies used this design since first appearing in the literature in the 1990s, with most published since the year 2000 (see Table 1).

Rubin (1991) aggregated a set of AB designs to evaluate outreach counseling and support groups for battered women. A formidable challenge to his evaluation effort was that interventions used at the agency were "somewhat unsystematic," theoretically eclectic, "idiosyncratic," and "... varied from session to session and across practitioners (group leaders)" (p. 340). With two sets of measures collected on each woman, one scoring desired thoughts, feelings, and behaviors and the other scoring abusive behavior by the spouse, neither measure showed positive change that could be attributed to agency interventions.

Johnson et al. (2001) obtained more favorable results when examining whether leader-

initiated changes influenced group processes in three AB design studies implemented by undergraduate students and practitioners. The three studies included (a) student leaders of an adolescent girls' group encouraging the development of trusting relationships between members and the effect on subsequent group attendance, (b) interventions to increase participation by the "silent majority" in a group of veterans and the number of members later speaking up in the group, and (c) encouragement of self-expression through puppetry in children who had experienced loss and emotional trauma and the number of children then using this mode of expression. Changes in dependent measures from the A to B phases were statistically significant in the first two studies. In the third study, statistical tests were not used, but there was 200% improvement from baseline to intervention mean scores.

Ledgerwood, Alessi, Hanson, Godley, and Petry (2008) used counterbalanced AB designs ( $A^1A^2B$ ;  $A^1BA^2$ ) across four clinics to gauge the effects of contingency management on attendance at substance abuse therapy groups. Three randomly selected clinics were exposed to an  $A^1A^2B$  design composed of the following conditions: An 8-week "baseline" of standard treatment ( $A^1$ ), 16 additional weeks of standard treatment ( $A^2$ ), and 16 weeks of standard treatment plus contingency management (material incentives for attendance) (B). A fourth randomly selected clinic was exposed to an  $A^1BA^2$  design, which resembled the previous design except that it presented the last two phases in reverse order. Clients who were enrolled in the clinics when contingency management was implemented attended a significantly greater percentage of therapy sessions than clients who were enrolled in treatment without contingency management.

### ABC, ABCD, $AB^1B^2$ , $AB^1B^2B^3$ , and Related Designs

If the outcome in the B phase of an AB design is unsatisfactory, a second intervention (C) can be tested in an ABC design, or the same intervention at a higher magnitude ( $B^2$ ) can be tested in an  $AB^1B^2$  design. If the outcome of C or  $B^2$  proves unsatisfactory, a third treatment (D) can be tested in an ABCD design, or the same treatment can be tested at an even higher

magnitude ( $B^3$ ) in an  $AB^1B^2B^3$  design. This highlights one of the advantages of SCDs; namely, their flexibility and ability to incorporate an evidence-based, trial-and-error process within clinical practice. This flexibility is crucial for accommodating the variable responses of individuals and groups to particular interventions.

Although the ABC, ABCD,  $AB^1B^2$ ,  $AB^1B^2B^3$ , and related designs permit the modification of interventions until the client or group has improved to a satisfactory level, these designs are unable to determine which of the interventions, if any, actually produced the improvement. This is because these designs are merely complicated versions of AB designs and are also susceptible to extraneous events that might have coincided with any of the seemingly effective treatment phases. In addition, sequence and carryover effects are possible in ABC and ABCD designs in that positive changes in C or D phases actually might be due to delayed effects of previous treatment phases. Interpretation of group work data is even trickier, because the conduct and affect of group members and the nature of the group experience is expected to change as a group moves through developmental stages.

Two studies have recently utilized these designs. Fry, Botterill, and Pring (2009) evaluated two formats of group-based speech-language therapy on the stuttering of an 18-year-old male client (in a group of up to 10 members) using an ABC design with a maintenance phase. After a 5-week baseline (A), 2 weeks of intensive group instruction was provided during one treatment phase (B), followed by 5 weeks of consolidation/self-management training in a second treatment phase (C). Finally, follow-up measures were taken during a 10-month maintenance phase. Compared to the A phase, there was a significant reduction in the percent of stuttered syllables per session and in the duration of stuttering in the B and C phases, and this improvement continued or increased in the maintenance phase. Although introduction of the interventions was associated with decreased stuttering, as intended, it was not known if the interventions were responsible for this improvement.

A group intervention has also been added to other therapeutic modalities in a stepwise ABC treatment design. Phaneuf and McIntyre (2011)

used such a design in a three-tier parent training program involving parents of preschool children with developmental disabilities ( $n = 8$  parent-child dyads). Each tier included an intervention component followed by an evaluation to determine who either responded favorably to the intervention, or who would need further service in the next intervention tier. The first phase included 3 weeks of individual participants listening to or reading materials drawn from Webster-Stratton's Incredible Years program, followed by an evaluation to determine either who responded and would be scheduled for a 3-month follow up assessment, or who would move on to the second intervention tier. The second tier involved 11 weeks of Webster-Stratton's group-based Incredible Years Parent Training Program, followed by another evaluation. Those who responded were scheduled for the 3-month follow up, and those who needed the final tier received a single session of individualized video feedback training and were scheduled for the last 3-month follow up. All but one child displayed reductions in observed inappropriate behavior or parental ratings of child behavior problems.

### **ABA, BAB, ABAB, ABCBCB, AB<sup>1</sup>B<sup>2</sup>B<sup>1</sup>B<sup>2</sup>, and Other Reversal Designs**

Inherent weaknesses of AB, ABC, and ABCD designs can be rectified by removing the seemingly effective treatment and returning to the previous experimental condition to reverse observed improvements, and then reintroducing the treatment to replicate its positive effects. Replication of treatment effects by reapplying an intervention and reproducing its results is a powerful demonstration of causality (Sidman, 1960). The drawbacks of reversal designs include ethical concerns about removing an apparently effective treatment merely for the purpose of demonstrating its efficacy, irreversible effects of certain interventions (such as psychoeducational therapies conveying knowledge that cannot be removed), and changes in certain group-based dependent variables (e.g., cohesion) attributable to group maturation that also might be irreversible. Ten reversal designs have been used in group work since the early 1970s, representing the second most utilized SCD (see Table 1).

In a landmark study, both for its use of precise observational measures of group interaction and SCD evaluation methodology, Linsk, Howe, and Pinkston (1975) included several ABAB designs to gauge the effect of question-asking on appropriate verbalizations by elderly residents in activity groups in a nursing home. Following a baseline phase in which group activities were conducted in the agency's usual manner, a second phase gauged the effects of the group leader asking residents as many questions as possible during the group activity. In a third phase, the group leader returned to the customary manner of directing the group. In the fourth and final treatment phase, the group leader resumed the procedure of increased question-asking. Linsk and colleagues showed that more frequent questions doubled the number of appropriate verbalizations relative to baseline sessions using the agency's usual procedure.

Weber (1980) investigated the impact of videotape feedback on adolescents in inpatient group therapy, hypothesizing that it would increase verbal expressions of warmth and decrease expressions of hostility and flight. During experimental sessions, the first 45 minutes of each group session was videotaped and then immediately played back to the members. The effects of videotaping and playback were evaluated within an ABAB design with each phase lasting 2 weeks. There were inconsistent results across the three measures of verbal affect. There was some support for an increase in warmth expressed during videotaping. However, trends within phases were unstable with warmth increasing in the initial baseline phase without videotape feedback and decreasing rapidly at the end of the final videotape feedback phase. For flight there was some reduction in the number of responses during videotape feedback; but, here too, flight responses were declining steadily at the end of both baseline phases without videotape feedback. Finally, hostility showed no consistent reduction during the presence of videotape feedback. Different statistical tests for change in the SCD data yielded conflicting results, showing how these analyses can be contradictory. Contributing to the inconsistent findings was high variability within the phases and a predetermined length for all phases (2 weeks). Sometimes experimental conditions were altered before data within each phase had stabilized. Had the investigator waited until data



within phases had stabilized before changing conditions, he might have been able to more clearly identify the levels and trends within phases and then been able to detect changes in subsequent phases.

Martinez and Wong (2009) utilized an ABAB design to evaluate the effects of written and telephone prompts on attendance at group support meetings for women survivors of domestic abuse. Participants of this study were residents of a transitional housing program who upon admission to the program were notified of the group support meetings along with other services offered by the agency. After 2 weeks of baseline in which no additional steps were taken to increase attendance at the support meetings, a colorful written invitation to attend the meeting was delivered and a single phone call was made to each resident the morning of the meeting for 2 weeks. Next, the baseline condition of no prompts was reinstated for 2 weeks, which was followed by a final 3-week phase in which written and telephone prompts were reintroduced. During baseline phases attendance was limited to the same three residents. During the first prompting phase the number of residents in attendance doubled. However, when prompts were discontinued attendance fell back to the level observed in baseline. In the final prompting phase the number of residents again increased, this time to more than triple what was recorded in baseline.

### Multiple-Baseline Designs (MBDs)

MBDs have been the most frequently used SCD in group work ( $k = 21$ ), probably because they lack troublesome reversal phases and because group work involves multiple clients whose progress sometimes can be monitored within separate baselines. Also contributing to the common use of this design is the prevalence of social skills training in group work where MBDs have been utilized to assess the effects of sequential training of specific social skills. This section will review several illustrative group work studies using concurrent and nonconcurrent MBDs.

**Concurrent MBD.** As noted above, ethical concerns and irreversibility of certain types of change are limitations of ABAB and other reversal designs. Concurrent MBDs do not contain reversal phases and circumvent these prob-

lems by simultaneously collecting data on two or more independent baselines—consisting either of different behaviors/targets, the same behaviors/targets in different settings, or the same behaviors/targets in different group members (or groups)—and then sequentially introducing the intervention across these baselines. If improvements occur when, and only when, an intervention is successively applied to two or more of these baselines, then the intervention's effectiveness is demonstrated. Because concurrent MBDs replicate treatment effects across behaviors or across group members, they can demonstrate some treatment impact across those two dimensions. Drawbacks to using concurrent MBDs can be the lack of two or more independent but equivalent baselines (which must respond similarly to the same intervention) and the requirement to delay the intervention for the second and subsequent baselines. The concurrent MBD was used in 17 group work studies.

Hansen, St. Lawrence, and Christoff (1985) delivered a problem-solving skills intervention for challenging situations in community living to seven patients in a partial hospitalization program. Hansen and colleagues utilized instructions, modeling, behavior rehearsal, and verbal reinforcement during group sessions to successively teach skill components of problem identification, goal definition, solution evaluation, evaluation of alternatives, and selection of the best alternative. Group data plotted within a concurrent MBD across skill components showed that the five skills were rarely exhibited during baseline phases, but these same skills showed immediate and significant increases following training.

Hall, Dineen, Schlesinger, and Stanton (2000) used a MBD to evaluate the effectiveness of group training in verbal interactive skills for developmentally disabled adults in a vocational training program. Six complex verbal skills (e.g., "social conversation") were successively taught with instructions, modeling, rehearsal, and feedback over a 7-week period, and the effects of training were evaluated with a concurrent MBD across responses. There were mixed or null effects of the training procedure on clients' rated skills. Hall and colleagues noted that this poor outcome might have been due to their training schedule that required leaders to begin training on additional skills before

previously trained skills had been mastered, and the low proficiency of group trainers who were mostly first-year graduate students.

**Nonconcurrent MBD (NCMBD).** NCMBDs are a variant of the MBD in which the baselines do not all start at the same time or run simultaneously. If implemented strictly according to the first published description of this design (Watson & Workman, 1981), the pre-treatment baseline phases are of different lengths and the assignment of these different baseline lengths to individual clients is randomized. In the published literature, however, NCMBDs containing different baseline lengths and random assignment of these baselines has been inconsistent. Because the separate baselines of a NCMBD are initiated at different times, and can even be started after the completion of other baselines, a NCMBD is essentially a series of AB designs and is relatively easy to carry out. Furthermore, because replication of intervention effects in the separate AB designs of the NCMBD are conducted with different participants, this design demonstrates some generality of intervention effects across subjects. There were four examples of this design in the literature (see Table 1).

Connors, Johnson, and Stuckey (1984) implemented an NCMBD in a study of brief psychoeducational group therapy for the treatment of bulimia. Treatment consisted of education, self-monitoring, goal setting, assertion training, relaxation, and cognitive restructuring, which was applied with two groups of 10 normal-weight bulimic women, one group 3 weeks before the other. The investigators reported a 70% reduction in binge/purge episodes and significant improvements in self-esteem, depression, assertiveness, and attitudes about eating that coincided with the introduction of treatment for each group.

Edleson, Miller, Stone, and Chapman (1985) used a “modified” MBD that resembled a NCMBD to evaluate a 12-week, small-group treatment for men who battered their partners. Groups were guided through topics of analyzing chains of events leading to violent situations, progressive relaxation, cognitive restructuring, problem solving, taking the partner’s view of the problem, and negotiation skills. Three groups began treatment at three different points in time. Edleson and colleagues reported substantial declines in the mean incidence of self-reported spouse abuse in two of the three groups as they began

participating in the program, which they pointed to as evidence of the program’s effectiveness. A couple of features of this study, however, limit its ability to demonstrate treatment efficacy: (a) the baseline data for all three groups were reconstructed retrospectively, which greatly increased the possibility of error; and (b) group treatment was preceded by 2 weeks of intake interviews, during which all groups showed immediate “decreases” in spouse abuse from baseline levels. The decline during the interview period, before any treatment was administered, suggests that decreases in abuse may have been due to the intake interviews or participants’ admission to the program rather than the treatment itself.

Kettlewell, Mizes, and Wasylshyn (1992) met all of the strict requirements for a NCMBD in a study evaluating group cognitive-behavioral treatment for bulimia. Kettlewell and his associates used a comprehensive treatment that included information on medical and psychological consequences of bulimia, goal setting, functional analysis of the binge-purge cycle, selecting an appropriate weight for oneself, coping strategies, combating irrational cognitions about thinness, and ways to maintain progress. This treatment was applied after 2, 6, or 10 weeks of stable baselines with three groups of clients, the length of baseline for each group being randomly determined. The weekly rate of bingeing and purging markedly declined for all three groups after the onset of treatment and, with the exception of one client, lower rates of eating disorders were mostly maintained during follow-up assessments.

### Alternating Treatments Design (ATD)

Group workers may want to know which intervention is most efficacious for their members, and the ATD (Barlow & Hayes, 1979) is a design specifically for the purpose of comparing two or more interventions. The ATD makes comparisons between interventions by rapidly alternating (e.g., from one group session to the next) between two or more treatment conditions and baseline conditions on a random schedule. Sometimes a sign or cue is used to identify the session (e.g., “Positive Interactions Session”) so that participants can distinguish which condition is currently in effect. The resulting data are graphed and data points from the same condi-

tion are connected by lines to form curves representing group members' performance during the different conditions. The distance separating the curves represents the differential effects of the various conditions. The ATD is an efficient design that can reveal differences between conditions after a small number of sessions. Potential problems of the ATD are contrast effects and multiple-treatment interference, which are experimental artifacts brought on by close juxtaposition of intervention conditions or cross-contamination and interaction of the different conditions, respectively. Given that groups develop and mature over time, there is a good likelihood of interdependency between phases, producing such interaction effects. These threats can be reduced by ending the ATD with a final phase in which the intervention with the best outcome is presented alone consecutively for several sessions. If performance within this final phase is similar to that recorded in the earlier alternating treatments phase of the study, then the results were not due to contrast effects or multiple-treatment interference.

As seen in Table 1, the first publications that included the ATD in group work appeared in the 1980s ( $k = 2$ ), with only two more over the next two decades. Wong and colleagues (1988) used an ATD to assess the effects of two types of structured group activities (e.g., team sports, art projects) on bizarre psychotic and appropriate behaviors in 10 male chronic psychiatric patients on a hospital ward. On a random schedule, the group of patients participated in one of three different types of sessions: (a) structured group activities with positive reinforcement (tokens earned through active participation), (b) structured group activities with noncontingent reinforcement (tokens given regardless of the level of participation), or (c) baseline sessions (control sessions in which no structured activity was offered). Psychotic behavior was significantly lower (70%) and appropriate behavior was substantially higher (approximately 500%) in the recreational sessions as opposed to the baseline sessions. Data for the two recreational conditions were roughly equivalent indicating that the beneficial effect was due to the recreational activities and not contingent reinforcement.

Blake, Owens, and Keane (1990) used an ATD to evaluate the efficacy of a single pack-

age intervention, verbal and written announcements of daily activities plus individual feedback and praise, on patients' attendance at group activities on a postacute psychiatric transition unit. On a weekly basis for the 20 weeks of the study, these procedures to encourage attendance alternated randomly with the unit's customary procedure of having nursing staff verbally remind patients that they were expected to attend all groups as part of their treatment. Attendance steadily increased in both conditions during the length of the study, but attendance was significantly higher in the prompting and feedback condition.

Kastner and May (2009) utilized an ATD to gauge the effects of action-oriented techniques (AO) on group climate (using the Group Climate Questionnaire—Short Version, MacKenzie, 1983) and disruptive out-of-group behavior on a group of seven middle-school students referred for behavioral or emotional difficulties. Group sessions consisted of discussion on topics selected by group members and alternated randomly between sessions with AO (i.e., the double, the empty chair, role reversal, and psychodramatic and behavioral role play) and sessions without AO. There were no significant differences between the two conditions in the frequency of disruptive behaviors, ratings of group engagement, or ratings of avoidance. However, ratings of engagement increased and ratings of avoidance decreased significantly over time, whereas ratings of conflict were unchanged.

### Mixed and Combined Designs

Although articles and textbooks on SCDs present these formats as relatively fixed structures, the essence of all of these designs is the systematic manipulation of the independent variable to produce a corresponding pattern of changes in the dependent variable that is unlikely to have been due to random fluctuation or extraneous events. Researchers who appreciate the underlying logic of SCDs can fully exploit their flexibility by mixing or combining the manipulations characteristic of the different designs (e.g., reversals, successive application of treatment across behaviors) to fit the needs of the particular study or to investigate additional questions of interest (Kazdin, 2011; Kennedy, 2005; Smith, 2012; Wong & Liberman, 1981).

The major drawback to mixed and combined designs is that they are more intricate and require more sophisticated understanding of SCD methodology.

Kohler and Fowler (1985) used a combination of concurrent MBDs and ABA designs in a study teaching three young children to invite peers to play and to use social amenities (e.g., “thank you,” “please”). The study evaluated effects of group training and the termination of training (during the reversal phase) on the children’s behavior and their peers’ responses outside of group in a classroom. For two of the three children, the MBD showed that training increased invitations and amenities performed by the children and those responses were reciprocated by their peers. Termination of training resulted in a reduction of invitations and amenities, but not to the near zero levels observed in the initial baseline. For the third child (who was only taught invitations), group training increased the rate of invitations in the target child but these responses were not reciprocated by her peers, and termination of training resulted in a return to the low levels of invitations observed in baseline. For this child, additional training of the target child and her peers along with group rewards was implemented to raise the number of invitations. This condition increased both the number of invitations given by the child and made by her peers, and when these procedures were terminated, the level of invitations remained above that recorded in previous baseline phases. The complicated questions about lasting effects of training and the testing of procedures to increase invitations for the third child could not have been answered by using a MBD or an ABA design alone. Additional experimental manipulations were needed for these purposes.

Wong, Morgan, Crowley and Baker (1996) used a mixed ABC design and a concurrent MBD across subjects to evaluate minimal in-situ training to promote generalization of social skills taught to three adolescent psychiatric inpatients. After collecting baseline data (A), the “Stacking the Deck” table game (Foxx, McMorow, & Schloss, 1983) (B) was used to teach and reinforce appropriate interactive skills to the three adolescents. The game involved responding to social scenarios described on game playing cards, modeling of correct responses, giving praise and the opportunity to move one’s game piece after correct responses to the situa-

tion, and positive (monetary) reinforcement. Covert generalization assessments conducted on the hospital ward revealed that despite substantial increases in trained responses in the game setting, these skills were rarely exhibited on the ward where patients resided. To promote generalization of the skills beyond the game situation, minimal in-situ training involving intermittent verbal prompts and tangible reinforcement (C) were administered when the skills were used during social interactions on the ward. Minimal in-situ training was successfully applied with each of the three adolescents immediately producing substantial improvement in skill usage on the ward in each patient, thereby replicating treatment effects in the pattern of a concurrent MBD across subjects. Thus, this study utilized an ABC design to show that game training (B) failed to produce changes that generalized to the more naturalistic ward setting, but that minimal in-situ training (C) did engender those changes, which was proven by replicating those improvements across subjects in the concurrent MBD.

### Discussion and Recommendations

Although group work researchers have not used SCDs frequently, they have used a range of designs to evaluate diverse interventions aimed at changing a broad spectrum of dependent variables. The number of published reports using SCDs in group work found in this review ( $k = 51$ ) is substantially larger than the two reported in Tolman and Molidor’s paper (1994). After a decrease in publications in the 1990s, there has been growth in the past decade. Perhaps this is an encouraging sign.

SCDs have been used to investigate both processes and outcomes of group therapy using both individual and group levels of measurement. Most studies focused on group outcomes, across a broad range of target areas, including social skills, eating disorders, anxiety, self-expression, warmth and hostility, and self-efficacy. The variables focusing on the group itself (e.g., processes) included attendance (Blake et al., 1990; Helmus, Saules, Schoener, & Roll, 2003; Johnson et al., 2001; Kirby, Kerwin, Carpenedo, Rosenwasser, & Gardner, 2008; Martinez & Wong, 2009; Petry et al., 2001), verbal participation (Gajar, Schloss, Schloss, & Thompson, 1984; Hauserman, Zwe-

back, & Plotkin, 1972; Johnson et al., 2001; Kohler & Fowler, 1985; Linsk et al., 1975; Tracey, Briddell, & Wilson, 1974), group interaction (Toseland, et al., 1978), language patterns (Fontao & Mergenthaler, 2008), group climate (Kastner & May, 2009), and satisfaction with group (O'Brien, et al., 1999; Patterson & Basham, 2002).

Most studies used the individual as the level of measurement and then derived a group mean score for analytical purposes. This method of analysis, though convenient and appropriate in many cases, obscures individual and subgroup differences. Patterson and Basham (2002) examined satisfaction with group at both the individual and group levels using an innovative mapping procedure, showing the importance of examining both simultaneously. Rather than using individual scores or the mean score for the group as the unit of analysis, this approach showed variability in both individual and group outcomes. Researchers should be aware of the benefits and limitations of individual and group levels of measurement and analysis, and select the most appropriate method.

The quality and rigor of group work studies using SCD methodology varied to a large degree. SCDs are merely evaluation tools and they cannot rise above other key components of the clinical study. For example, in the case of Rubin (1991), SCDs were unable to produce consistent or interpretable results for interventions that the author himself admitted were unsystematic and vacillated from session to session. It is unclear how such a study could yield reliable results, or if such results emerged how they would be understood. This also raises the question about what treatments lend themselves to SCDs. As indicated in Table 1, most interventions incorporated forms of behavior therapy, but others utilized solution focused (Nelson & Kelley, 2001), experiential (Patterson & Basham, 2002), couples group therapy (Nelson & Kelley, 2001), and other well-defined techniques (Johnson et al., 2001). This shows that SCDs can accommodate many different types of therapy. However, whatever theoretical treatment approach is being evaluated, it must be applied systematically and consistently to produce interpretable results. This is true of SCDs as well as between-groups designs.

The study by Edleson and colleagues (1985) presented a different set of concerns by using a

reconstructed baseline and interpreting the resulting data in a questionable manner. The doubtful accuracy of reconstructed baseline data was a weakness in itself, but that combined with desired changes occurring prior to the introduction of therapy should have discouraged the authors from drawing conclusions about treatment efficacy. Most of the other studies cited here, however, applied SCDs in a technically sound manner and their results were more definitive.

### Recommendations for Research

Use of SCDs in group work appears to be at an intermediate level of development where a number of designs have been used and some fairly often, but there are opportunities for more varied and refined applications. Group work researchers have not always distinguished between uncontrolled case study SCDs (e.g., B, AB, ABC) and controlled experimental SCDs (e.g., ABAB, MBDs), only the latter of which replicate treatment effects and thereby permit logical inference of treatment efficacy. There are efforts underway to raise the standards for SCD research. Guidelines have been developed for rating SCDs based on their level of evidence for demonstrating a causal relationship (Smith, 2012). According to such guidelines (Romeiser Logan, Hickman, Harris, & Heriza, 2008; What Works Clearinghouse, 2013), a MBD with three baselines could provide a convincing demonstration of treatment efficacy, but a MBD with less than three baselines could not. This is noteworthy, as the proposed standard is more rigorous and stringent than that previously presented in published research and textbooks covering SCDs (Bailey & Burch, 2002; Cooper, Heron, & Heward, 2007).

Future research could apply SCDs not used in previous group work studies. One example is the Changing Criterion Design, which is applicable in situations where a dependent variable can be tightly controlled (Hall & Fox, 1977; Hartmann & Hall, 1976). This design requires one group member or group and one independent variable that produces stepwise changes in the dependent variable. For example, it may be unreasonable to expect that a group member with limited social competence be fully engaged in the group within a few sessions as a result of an engagement-building intervention. Instead, incremental engagement objectives

may be set over a number of group sessions (e.g., 25% more engaged per set of sessions), leading to full engagement as the ultimate goal. Other novel designs and experimental manipulations that could be used to explore group work questions or to strengthen the internal validity of a SCD study involve randomization. Points in time that interventions are started or the order in which group members or groups are exposed to treatment are randomly determined, and such randomization can be incorporated into almost any SCD (Edgington, 1975; Kratochwill & Levin, 2010). Randomization is likely to require SCDs with longer baseline phases, but this may be acceptable in some group work research (see, for example, Toseland, et al., 1978).

An important question that group workers could explore using SCDs is whether group processes directly impact group outcomes. This research could be conducted with two-stage SCDs where both processes and outcomes are simultaneously monitored, and data are analyzed to see if changes in relevant processes reliably precede changes in desired outcomes. This type of investigation would be difficult to conduct using conventional between-groups research designs, but would be manageable using within-groups SCDs. There are many measures of group processes, such as group climate (MacKenzie, 1983, as used in Kastner & May, 2009) and group engagement (Macgowan & Newman, 2005), but the most frequently studied variable has been cohesion, which has been reliably associated with outcomes in many group studies (Burlingame, McClendon, & Alonso, 2011; Burlingame et al., 2006; Strauss, Burlingame, & Bormann, 2008). As Fike (1980) argued, it is important to capture not only individual outcomes, but also factors that reflect the group experience on individuals and on the group, such as satisfaction and the nature of interactions.

### Recommendations for Practice

The goals, resources, and constraints for practice differ substantially from that for research, and recommendations for SCDs for practice should reflect those differences (Wong, 2010). Recall that our review of group research from 1972 to 2012 found an average of only slightly more than one SCD study published per year. Guidelines for SCD research calling for

more rigorous designs with stronger evidence for causality could have a suppressive effect on the usage of these designs in evaluating practice. Standards that devalue weaker SCDs might result in even fewer SCDs being applied and less attempts to empirically evaluate group work. Individual evaluation, even if imperfect, is preferable to automatically assuming that a treatment validated in a prior between-groups study will be equally effective with the group of clients one is currently working. It is also clearly superior to practice relying on tradition, popularity of techniques, personal belief, or other approaches not based on client data.

There has been much discussion about the importance of developing and/or adapting existing treatments to maximize efficacy with particular client populations (Benish, Quintana, & Wampold, 2011; Castro, Barrera, & Steiker, 2010; Chen, Kakkad, & Balzano, 2008). La Roche and Christopher (2008) noted that no empirically supported treatment has met all of the criteria set forth by the APA Task Force for treatment efficacy (American Psychological Association, 2006) for any minority group. In adapting existing group treatments, components are added that attend to culture, while retaining the “essential” ingredients of the established group treatment (Macgowan & Hanbidge, 2014). A MBD across different groups or settings could be used to test the relative effectiveness of a group intervention that is progressively adapted to include important cultural components. If the SCD is done rigorously (Chambless et al., 1998; Kratochwill & Levin, 2010), these data can contribute to the development of an empirically supported group intervention appropriate for that population.

Uncontrolled case study designs such as the B, AB, and ABC designs should be used in clinical practice, but with full awareness of their limited inferential power. Such designs do not provide the controls or replications necessary to allow clinicians to logically conclude that their interventions caused whatever changes might have occurred. Nevertheless, these designs are valuable because “even if the therapist and clients are not certain that the intervention was the sole cause of the changes, it still makes it possible to ascertain the degree to which goals are being achieved” (Rose, 1977, p. 67).

Group work practitioners need not be constrained by earlier group work evaluative meth-

ods. They can apply more refined variants of the B design. Examples include the Periodic Treatments Design (Hayes, 1981; Hayes, Barlow, & Nelson-Gray, 1999) and the Repeated Pretest-Posttest Design (Thyer & Curtis, 1983), which are controlled designs that can evaluate an intervention during a single intervention phase that is neither preceded nor followed by a baseline phase. These designs can demonstrate experimental control by replicating improvements occurring soon after periodic exposure to intervention (e.g., a technique to improve group cohesion) while showing the absence of such improvements at other times during the phase. In addition, the internal validity of other case study designs, such as the AB or ABC, can be strengthened with the earlier mentioned procedure of randomly selecting the starting point for interventions (Edgington, 1975; Kratochwill & Levin, 2010; Toseland, et al., 1978).

### Conclusions

This paper has elucidated how SCDs have been used in group work by reviewing the designs in published studies, discussing their respective strengths and limitations, and examining procedural variations producing data that were relatively clear or confounded. Most of the group work studies reviewed here successfully used SCDs to accomplish their purpose—to systematically evaluate the effects of their interventions on processes and outcomes. Group researchers can learn from the mistakes and build upon the strengths of these earlier studies to become more facile with this research methodology. Although SCD technology has been used in group work since the early 1970s, there are many important questions that still need to be investigated and group researchers have only begun to tap the potential of these evaluative formats.

### References

- AASWG. (2006). *Standards for social work practice with groups* (2nd ed.). Retrieved from [http://www.aaswg.org/files/AASWG\\_Standards\\_for\\_Social\\_Work\\_Practice\\_with\\_Groups.pdf](http://www.aaswg.org/files/AASWG_Standards_for_Social_Work_Practice_with_Groups.pdf)
- AGPA. (2007). *Practice guidelines for group psychotherapy*. Retrieved September 5, 2013, from <http://www.agpa.org/guidelines/index.html>
- American Psychological Association. (2006). Evidence-based practice in psychology: APA Presidential Task Force on Evidence-Based Practice. *American Psychologist*, *61*, 271–285. doi:10.1037/0003-066X.61.4.271
- ASGW. (2000). *Professional standards for the training of group workers*. Retrieved September 6, 2013, from [http://www.asgw.org/training\\_standards.htm](http://www.asgw.org/training_standards.htm)
- ASGW. (2007). *Association for Specialists in Group Work: Best practice guidelines 2007 revisions*. Retrieved August 6, 2013, from [http://www.asgw.org/PDF/Best\\_Practices.pdf](http://www.asgw.org/PDF/Best_Practices.pdf)
- Bailey, J. S., & Burch, M. R. (2002). *Research methods in applied behavior analysis*. Thousand Oaks, CA: Sage.
- Barlow, A., Burlingame, G. M., & Fuhriman, A. (2000). Therapeutic application of groups from Pratt's "thought control classes" to modern group psychotherapy. *Group Dynamics: Theory, Research, and Practice*, *4*, 115–134. doi:10.1037/1089-2699.4.1.115
- Barlow, D. H., & Hayes, S. C. (1979). Alternating treatments design: One strategy for comparing the effects of two treatments in a single subject. *Journal of Applied Behavior Analysis*, *12*, 199–210. doi:10.1901/jaba.1979.12-199
- Barlow, S. H., Fuhriman, A., & Burlingame, G. M. (2004). The history of group counseling and psychotherapy. In J. L. DeLucia-Waack, D. A. Gentry, C. R. Kalodner, & M. Riva (Eds.), *Handbook of group counseling and psychotherapy* (pp. 3–22). Thousand Oaks, CA: Sage. doi:10.4135/9781452229683.n1
- Bates, P. (1980). The effectiveness of interpersonal skills training on the social skill acquisition of moderately and mildly retarded adults. *Journal of Applied Behavior Analysis*, *13*, 237–248. doi:10.1901/jaba.1980.13-237
- Benish, S. G., Quintana, S., & Wampold, B. E. (2011). Culturally adapted psychotherapy and the legitimacy of myth: A direct-comparison meta-analysis. *Journal of Counseling Psychology*, *58*, 279–289. doi:10.1037/a0023626
- Blake, D. D., Owens, M. D., & Keane, T. M. (1990). Increasing group attendance on a psychiatric unit: An alternating treatments design comparison. *Journal of Behavior Therapy and Experimental Psychiatry*, *21*, 15–20. doi:10.1016/0005-7916(90)90044-L
- Bloom, M., Fischer, J., & Orme, J. G. (2009). *Evaluating practice: Guidelines for the accountable professional* (6th ed.). Boston, MA: Pearson Allyn & Bacon.
- Burlingame, G. M., Strauss, B., & Joyce, A. (2013). Change mechanisms and effectiveness of small group treatments. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed., pp. 640–689). New York: Wiley.

- Burlingame, G. M., McClendon, D. T., & Alonso, J. (2011). Cohesion in group therapy. *Psychotherapy, 48*, 34–42. doi:10.1037/a0022063
- Burlingame, G. M., Strauss, B., Joyce, A., MacNair-Semands, R., MacKenzie, K. R., Ogrodniczuk, J., & Taylor, S. (2006). *CORE Battery—Revised: An assessment toolkit for promoting optimal group selection, process, and outcome*. New York, NY: American Group Psychotherapy Association.
- Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Boston, MA: Houghton Mifflin Company.
- Castro, F. G., Barrera, M., & Steiker, L. K. H. (2010). Issues and challenges in the design of culturally adapted evidence-based interventions. *Annual Review of Clinical Psychology, 6*, 213–239. doi:10.1146/annurev-clinpsy-033109-132032
- Cetingok, M., & Hirayama, H. (1983). Evaluating the effects of group work with the elderly: An experiment using a single-subject design. *Small Group Behavior, 14*, 327–335. doi:10.1177/104649648301400304
- Chambless, D. L., Baker, M. J., Baucom, D. H., Beutler, L. E., Calhoun, K. S., Crits-Christoph, P., . . . Woody, S. R. (1998). Update on empirically validated therapies, II. *The Clinical Psychologist, 51*, 3–16.
- Chen, E. C., Kakkad, D., & Balzano, J. (2008). Multicultural competence and evidence-based practice in group therapy. *Journal of Clinical Psychology, 64*, 1261–1278. doi:10.1002/jclp.20533
- Clark, M. L., Cunningham, L. J., & Cunningham, C. E. (1989). Improving the social behavior of siblings of autistic children using a group problem solving approach. *Child & Family Behavior Therapy, 11*, 19–33. doi:10.1300/J019v11n01\_02
- Connors, M. E., Johnson, C. L., & Stuckey, M. K. (1984). Treatment of bulimia with brief psychoeducational group therapy. *The American Journal of Psychiatry, 141*, 1512–1516.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Chicago, IL: Rand McNally College Pub. Co.
- Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Upper Saddle River, N. J.: Pearson/Merrill-Prentice Hall.
- Edgington, E. S. (1975). Randomization tests for one-subject operant experiments. *The Journal of Psychology: Interdisciplinary and Applied, 90*, 57–68. doi:10.1080/00223980.1975.9923926
- Edleson, J. L., Miller, D. M., Stone, G. W., & Chapman, D. G. (1985). Group treatment for men who batter. *Social Work Research & Abstracts, 21*, 18–21. doi:10.1093/swra/21.3.18
- Feldman, R. A. (1987). Group work knowledge and research: A two-decade comparison. *Social Work with Groups, 9*, 7–14. doi:10.1300/J009v09n03\_03
- Fike, D. F. (1980). Evaluating group intervention. *Social Work with Groups, 3*, 41–51. doi:10.1300/J009v03n02\_04
- Fontao, M. I., & Mergenthaler, E. (2008). Therapeutic factors and language patterns in group therapy application of computer-assisted text analysis to the examination of microprocesses in group therapy: Preliminary findings. *Psychotherapy Research, 18*, 345–354. doi:10.1080/10503300701576352
- Foster, L. H., Watson, T. S., & Young, J. S. (2002). Single-subject research design for school counselors: Becoming an applied researcher. *Professional School Counseling, 6*, 146–154.
- Foxx, R. M., McMorrow, M. J., Bittle, R. G., & Ness, J. (1986). An analysis of social skills generalization in two natural settings. *Journal of Applied Behavior Analysis, 19*, 299–305. doi:10.1901/jaba.1986.19-299
- Foxx, R. M., McMorrow, M. J., & Schloss, C. N. (1983). Stacking the deck: Teaching social skills to retarded adults with a modified table game. *Journal of Applied Behavior Analysis, 16*, 157–170. doi:10.1901/jaba.1983.16-157
- Frank, J. D. (1975). Group psychotherapy research 25 years later. *International Journal of Group Psychotherapy, 25*, 159–162.
- Fry, J. P., Botterill, W. M., & Pring, T. R. (2009). The effect of an intensive group therapy program for young adults who stutter: A single subject study. *International Journal of Speech-Language Pathology, 11*, 12–19. doi:10.1080/17549500802600990
- Gajar, A., Schloss, P. J., Schloss, C. N., & Thompson, C. K. (1984). Effects of feedback and self-monitoring on head trauma youths' conversation skills. *Journal of Applied Behavior Analysis, 17*, 353–358. doi:10.1901/jaba.1984.17-353
- Galassi, J. P., & Gersh, T. L. (1993). Myths, misconceptions, and missed opportunity: Single-case designs and counseling psychology. *Journal of Counseling Psychology, 40*, 525–531. doi:10.1037/0022-0167.40.4.525
- Garrett, K. J. (2005). School social workers' evaluation of group work practices. *Children & Schools, 27*, 247–252. doi:10.1093/cs/27.4.247
- Gronna, S. S., Serna, L. A., Kennedy, C. H., & Prater, M. A. (1999). Promoting generalized social interactions using puppets and script training in an integrated preschool. A single-case study using multiple baseline design. *Behavior Modification, 23*, 419–440. doi:10.1177/0145445599233005
- Hall, J. A., Dineen, J. P., Schlesinger, D. J., & Stanton, R. (2000). Advanced group treatment for developmentally disabled adults with social skill deficits. *Research on Social Work Practice, 10*, 301–326.
- Hall, J. A., Schlesinger, D. J., & Dineen, J. P. (1997). Social skills training in groups with developmen-



- tally disabled adults. *Research on Social Work Practice*, 7, 187–201.
- Hall, K. R. (2006). Using problem-based learning with victims of bullying behavior. *Professional School Counseling*, 9, 231–237.
- Hall, R. V., & Fox, R. G. (1977). Changing-criterion designs: An alternate applied behavior analysis procedure. In B. Etzel, J. Leblanc, & D. M. Baer (Eds.), *New developments in behavioral research: Theory, method, and application* (pp. 151–166). New York, NY: Wiley.
- Hansen, D. J., St. Lawrence, J. S., & Christoff, K. A. (1985). Effects of interpersonal problem-solving training with chronic aftercare patients on problem-solving component skills and effectiveness of solutions. *Journal of Consulting and Clinical Psychology*, 53, 167–174. doi:10.1037/0022-006X.53.2.167
- Hansen, D. J., St. Lawrence, J. S., & Christoff, K. A. (1989). Group conversational-skills training with inpatient children and adolescents: Social validation, generalization, and maintenance. *Behavior Modification*, 13, 4–31. doi:10.1177/01454455890131001
- Hartmann, D. P., & Hall, R. V. (1976). The changing criterion design. *Journal of Applied Behavior Analysis*, 9, 527–532. doi:10.1901/jaba.1976.9.527
- Hauserman, N., Zwebach, S., & Plotkin, A. (1972). Use of concrete reinforcement to facilitate verbal initiations in adolescent group therapy. *Journal of Consulting and Clinical Psychology*, 38, 90–96. doi:10.1037/h0032410
- Hayes, S. C. (1981). Single case experimental design and empirical clinical practice. *Journal of Consulting and Clinical Psychology*, 49, 193–211. doi:10.1037/0022-006X.49.2.193
- Hayes, S. C., Barlow, D. H., & Nelson-Gray, R. O. (1999). *The scientist practitioner: Research and accountability in the age of managed care* (2nd ed.). Boston, MA: Allyn & Bacon.
- Helmus, T. C., Saules, K. K., Schoener, E. P., & Roll, J. M. (2003). Reinforcement of counseling attendance and alcohol abstinence in a community-based dual-diagnosis treatment program: A feasibility study. *Psychology of Addictive Behaviors*, 17, 249–251. doi:10.1037/0893-164X.17.3.249
- Heppner, P. P., Kivlighan, D. M., & Wampold, B. E. (2008). *Research design in counseling* (3rd ed.). Belmont, CA: Thomson Brooks/Cole.
- Hersen, M., & Barlow, D. H. (1976). *Single case experimental designs: Strategies for studying behavior change*. New York, NY: Pergamon Press.
- Horne, A. M., & Rosenthal, R. N. (1997). Research in group work: How did we get where we are? *Journal for Specialists in Group Work*, 22, 228–240. doi:10.1080/01933929708415527
- Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, 71, 165–179.
- Johnson, P., Beckerman, A., & Auerbach, C. (2001). Researching our own practice: Single system design for groupwork. *Groupwork*, 13, 57–72.
- Kastner, J. W., & May, W. (2009). Action-oriented techniques in adolescent group therapy. *Group*, 33, 315–327.
- Kazdin, A. E. (2003). *Research design in clinical psychology* (4th ed.). Boston, MA: Allyn & Bacon.
- Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings* (2nd ed.). New York, NY: Oxford University Press.
- Kelly, J. A., Laughlin, C., Claiborne, M., & Patterson, J. (1979). A group procedure for teaching job interviewing skills to formerly hospitalized psychiatric patients. *Behavior Therapy*, 10, 299–310. doi:10.1016/S0005-7894(79)80020-9
- Kelly, J. A., Wildman, B. G., Urey, J. R., & Thurman, C. (1980). Group skills training to increase the conversational repertoire of retarded adolescents. *Child & Family Behavior Therapy*, 1, 323–336. doi:10.1300/J473v01n04\_02
- Kennedy, C. H. (2005). *Single-case designs for educational research*. Boston, MA: Pearson/A & B.
- Kettlewell, P. W., Mizes, J. S., & Wasylshyn, N. A. (1992). A cognitive-behavioral group treatment of bulimia. *Behavior Therapy*, 23, 657–670. doi:10.1016/S0005-7894(05)80227-8
- Kirby, K. C., Kerwin, M. E., Carpenedo, C. M., Rosenwasser, B. J., & Gardner, R. S. (2008). Interdependent group contingency management for cocaine-dependent methadone maintenance patients. *Journal of Applied Behavior Analysis*, 41, 579–595. doi:10.1901/jaba.2008.41-579
- Kohler, F. W., & Fowler, S. A. (1985). Training prosocial behaviors to young children: An analysis of reciprocity with untrained peers. *Journal of Applied Behavior Analysis*, 18, 187–200. doi:10.1901/jaba.1985.18-187
- Kratochwill, T. R., & Levin, J. R. (2010). Enhancing the scientific credibility of single-case intervention research: Randomization to the rescue. *Psychological Methods*, 15, 124–144. doi:10.1037/a0017736
- La Roche, M., & Christopher, M. S. (2008). Culture and empirically supported treatments: On the road to a collision? *Culture & Psychology*, 14, 333–356. doi:10.1177/1354067X08092637
- Ledgerwood, D. M., Alessi, S. M., Hanson, T., Godley, M. D., & Petry, N. M. (2008). Contingency management for attendance to group substance abuse treatment administered by clinicians in community clinics. *Journal of Applied Behavior Analysis*, 41, 517–526. doi:10.1901/jaba.2008.41-517
- Lees, D. G., & Ronan, K. R. (2008). Engagement and effectiveness of parent management training (incredible years) for solo high-risk mothers: A mul-

- multiple baseline evaluation. *Behaviour Change*, 25, 109–128. doi:10.1375/behc.25.2.109
- Linsk, N., Howe, M. W., & Pinkston, E. M. (1975). Behavioral group work in a home for the aged. *Social Work*, 20, 454–463.
- Lumpkin, P. W., Silverman, W. K., Weems, C. F., Markham, M. R., & Kurtines, W. M. (2002). Treating a heterogeneous set of anxiety disorders in youths with group cognitive behavioral therapy: A partially nonconcurrent multiple-baseline evaluation. *Behavior Therapy*, 33, 163–177. doi:10.1016/S0005-7894(02)80011-9
- Lundervold, D. A., & Belwood, M. F. (2000). The best kept secret in counseling: Single-case ( $N=1$ ) experimental designs. *Journal of Counseling & Development*, 78, 92–102. doi:10.1002/j.1556-6676.2000.tb02565.x
- Macgowan, M. J. (2008). *A guide to evidence-based group work*. New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780195183450.001.0001
- Macgowan, M. J., & Hanbidge, A. S. (2014). Advancing evidence-based group work in community settings: Methods, opportunities, and challenges. In J. L. DeLucia-Waack, C. R. Kalodner, & M. Riva (Eds.), *The handbook of group counseling and psychotherapy* (2nd ed., pp. 303–317). Thousand Oaks, CA: Sage.
- Macgowan, M. J., & Newman, F. L. (2005). The factor structure of the Group Engagement Measure. *Social Work Research*, 29, 107–118. doi:10.1093/swr/29.2.107
- MacKenzie, K. R. (1983). The clinical application of a group climate measure. In R. R. Dies & K. R. MacKenzie (Eds.), *Advances in group psychotherapy: Integrating research and practice* (pp. 159–170). New York, NY: International Universities Press.
- Marr, D. D., & Fairchild, T. N. (1993). A problem-solving strategy and self-esteem in recovering chemically dependent women. *Alcoholism Treatment Quarterly*, 10, 171–186. doi:10.1300/J020V10N01\_11
- Martinez, K. K., & Wong, S. E. (2009). Using prompts to increase attendance at groups for survivors of domestic violence. *Research on Social Work Practice*, 19, 460–463. doi:10.1177/1049731508329384
- Maughan, D. R., Christiansen, E., Jenson, W. R., Olympia, D., & Clark, E. (2005). Behavioral parent training as a treatment for externalizing behaviors and disruptive behavior disorders: A meta-analysis. *School Psychology Review*, 34, 267–286.
- McGrath, J. E. (1997). Small group research: That once and future field: An interpretation of the past with an eye to the future. *Group Dynamics: Theory, Research, and Practice*, 1, 7–27. doi:10.1037/1089-2699.1.1.7
- Mouton, S. G., & Stanley, M. A. (1996). Habit reversal training for trichotillomania: A group approach. *Cognitive and Behavioral Practice*, 3, 159–182. doi:10.1016/S1077-7229(96)80036-8
- Nelson, T. S., & Kelley, L. (2001). Solution-focused couples group. *Journal of Systemic Therapies*, 20, 47–66. doi:10.1521/jsyt.20.4.47.23085
- Newman, F. L., & Wong, S. E. (2004). Progress and outcome assessment of individual patient data: Selecting single subject design and statistical procedures. In M. E. Maruish (Ed.), *Use of psychological testing for treatment planning & outcome assessment* (3rd ed., Vol. 1, pp. 273–289). Mahwah, NJ: Erlbaum.
- O'Brien, W. H., Korchynsky, R., Fabrizio, J., McGrath, J., & Swank, A. (1999). Evaluating group process in a stress management intervention: Relationships between perceived process and cardiovascular reactivity to stress. *Research on Social Work Practice*, 9, 608–630. doi:10.1177/104973159900900506
- Patterson, D. A., & Basham, R. E. (2002). A data visualization procedure for the evaluation of group treatment outcomes across units of analysis. *Small Group Research*, 33, 209–232. doi:10.1177/104649640203300203
- Petry, N. M., Martin, B., & Finocche, C. (2001). Contingency management in group treatment: A demonstration project in an HIV drop-in center. *Journal of Substance Abuse Treatment*, 21, 89–96. doi:10.1016/S0740-5472(01)00184-2
- Petry, N. M., Weinstock, J., & Alessi, S. M. (2011). A randomized trial of contingency management delivered in the context of group counseling. *Journal of Consulting and Clinical Psychology*, 79, 686–696. doi:10.1037/a0024813
- Phaneuf, L., & McIntyre, L. L. (2007). Effects of individualized video feedback combined with group parent training on maternal inappropriate behavior. *Journal of Applied Behavior Analysis*, 40, 737–741.
- Phaneuf, L., & McIntyre, L. L. (2011). The application of a three-tier model of intervention to parent training. *Journal of Positive Behavior Interventions*, 13, 198–207. doi:10.1177/1098300711405337
- Robison, F. F., Morran, D. K., & Hulse-Killacky, D. (1989). Single-subject research designs for group counselors studying their own groups. *Journal for Specialists in Group Work*, 14, 93–97. doi:10.1080/01933928908411892
- Roessler, R. T., & Lewis, F. D. (1984). Conversation skill training with mentally retarded and learning disabled sheltered workshop clients. *Rehabilitation Counseling Bulletin*, 27, 161–171.
- Romeiser Logan, L., Hickman, R. R., Harris, S. R., & Heriza, C. B. (2008). Single-subject research design: Recommendations for levels of evidence and

- quality rating. *Developmental Medicine & Child Neurology*, 50, 99–103. doi:10.1111/j.1469-8749.2007.02005.x
- Rose, S. D. (1977). *Group therapy: A behavioral approach*. Englewood Cliffs, NJ: Prentice Hall.
- Rubin, A. (1991). The effectiveness of outreach counseling and support groups for battered women: A preliminary evaluation. *Research on Social Work Practice*, 1, 332–357. doi:10.1177/104973159100100401
- Serna, L. A., Schumaker, J. B., Sherman, J. A., & Sheldon, J. B. (1991). In-home generalization of social interactions in families of adolescents with behavior problems. *Journal of Applied Behavior Analysis*, 24, 733–746. doi:10.1901/jaba.1991.24-733
- Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology*. New York, NY: Basic Books.
- Smith, J. D. (2012). Single-case experimental designs: A systematic review of published research and current standards. *Psychological Methods*, 17, 510–550. doi:10.1037/a0029312
- Stevens, R. E. S. (2006). Research ‘Alive’ in the social work classroom: Teaching graduate social work students to evaluate practice effectiveness. *Arete*, 29, 62–68.
- Strauss, B., Burlingame, G. M., & Bormann, B. (2008). Using the CORE-R battery in group psychotherapy. *Journal of Clinical Psychology*, 64, 1225–1237. doi:10.1002/jclp.20535
- Thyer, B. A., & Curtis, G. C. (1983). The repeated pretest-posttest single-subject experiment: A new design for empirical clinical practice. *Journal of Behavior Therapy and Experimental Psychiatry*, 14, 311–315. doi:10.1016/0005-7916(83)90073-3
- Tolman, R. M., & Molitor, C. E. (1994). A decade of social group work research: Trends in methodology, theory, and program development. *Research on Social Work Practice*, 4, 142–159. doi:10.1177/104973159400400202
- Toseland, R., Krebs, A., & Vahsen, J. (1978). Changing group interaction patterns. *Journal of Social Service Research*, 2, 219–232.
- Toseland, R. W., & Rivas, R. F. (2012). *An introduction to group work practice* (7th ed.). Boston, MA: Allyn & Bacon.
- Tracey, D. A., Briddell, D. W., & Wilson, G. T. (1974). Generalization of verbal conditioning to verbal and nonverbal behavior: Group therapy with chronic psychiatric patients. *Journal of Applied Behavior Analysis*, 7, 391–402. doi:10.1901/jaba.1974.7-391
- Tymchuk, A. J., Andron, L., & Rahbar, B. (1988). Effective decision-making/problem-solving training with mothers who have mental retardation. *American Journal on Mental Retardation*, 92, 510–516.
- Vaccaro, F. J. (1990). Application of social skills training in a group of institutionalized aggressive elderly subjects. *Psychology and Aging*, 5, 369–378. doi:10.1037/0882-7974.5.3.369
- Ventimiglia, J. A., Marschke, J., Carmichael, P., & Loew, R. (2000). How do clinicians evaluate their practice effectiveness? A survey of clinical social workers. *Smith College Studies in Social Work*, 70, 287–306. doi:10.1080/00377310009517593
- Vonk, M. E., Tripodi, T., & Epstein, I. (2007). *Research techniques for clinical social workers* (2nd ed.). New York, NY: Columbia University Press.
- Wade, C. M., Ortiz, C., & Gorman, B. S. (2007). Two-session group parent training for bedtime noncompliance in Head Start preschoolers. *Child & Family Behavior Therapy*, 29, 23–55. doi:10.1300/J019v29n03\_03
- Watson, P. J., & Workman, E. A. (1981). The non-concurrent multiple baseline across-individuals design: An extension of the traditional multiple baseline design. *Journal of Behavior Therapy and Experimental Psychiatry*, 12, 257–259. doi:10.1016/0005-7916(81)90055-0
- Weber, L. A. (1980). The effect of videotape and playback on an in-patient adolescent group. *International Journal of Group Psychotherapy*, 30, 213–227.
- What Works Clearinghouse. (2013). *Procedures and standards handbook* (Version 3.0). Retrieved from [http://ies.ed.gov/ncee/wwc/pdf/reference\\_resources/wwc\\_procedures\\_v3\\_0\\_draft\\_standards\\_handbook.pdf](http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v3_0_draft_standards_handbook.pdf)
- Wong, S. E. (2010). Single-case evaluation designs for practitioners. *Journal of Social Service Research*, 36, 248–259. doi:10.1080/01488371003707654
- Wong, S. E., & Liberman, R. P. (1981). Mixed single-subject designs in clinical research: Variations of the multiple-baseline. *Behavioral Assessment*, 3, 297–306.
- Wong, S. E., Morgan, C., Crowley, R., & Baker, J. N. (1996). Using a table game to teach social skills to adolescent psychiatric inpatients: Do the skills generalize? *Child & Family Behavior Therapy*, 18, 1–17. doi:10.1300/J019v18n04\_01
- Wong, S. E., Wright, J., Terranova, M. D., Bowen, L., Zarate, R., & Zarate, R. (1988). Effects of structured ward activities on appropriate and psychotic behavior of chronic psychiatric patients. *Behavioral Residential Treatment*, 3, 41–50. doi:10.1002/bin.2360030104
- Zastrow, C. (2006). *Social work with groups: A comprehensive workbook* (6th ed.). Belmont, CA: Thomson.

Received February 15, 2013

Revision received November 4, 2013

Accepted November 5, 2013 ■

## Team Task Conflict Resolution: An Examination of Its Linkages to Team Personality Composition and Team Effectiveness Outcomes

Thomas A. O’Neill  
University of Calgary

Natalie J. Allen  
Western University

By drawing attention to task conflict resolution and reporting on its empirical relations with key team variables, the present study offers a new perspective on team effectiveness. Specifically, we examined how three “dark” personality traits (Manipulativeness, Narcissism, and Secondary Psychopathy) relate to team conflict resolution, team innovation, and team task performance in student engineering design project teams. Interestingly, results indicate that mean team levels of Secondary Psychopathy is the most important predictor, and task conflict resolution is a mediator linking mean Secondary Psychopathy to team task performance. Furthermore, the prediction of team performance by Secondary Psychopathy is as strong as is any of the “Big 5” personality variables investigated in meta-analyses.

*Keywords:* team conflict, task conflict, task conflict resolution, team effectiveness, Dark Triad, personality, psychopathy

Work teams form the essential building blocks of modern organizations (Hackman, 2002). As noted by Kozlowski and Bell (2003), the adoption of teams emerged, in large part, from global competition and an emphasis on innovation, which created a need for organizations to be increasingly flexible and adaptable. Identifying the determinants of work team processes, innovation, and performance, therefore, is fundamental to our understanding of organizational effectiveness (see Sundstrom, McIntyre, Halfhill, & Richards, 2000).

Perhaps not surprisingly, team conflict is one of the most researched team processes (see Korsgaard, Jeong, Mahony, & Pitariu, 2008). Consider the recent meta-analysis by de Wit, Greer, and Jehn (2012); in that research, the authors report on data from 116 studies involving almost 9,000 teams. Many of the studies

place emphasis on task conflict, which is the extent to which team members have different opinions, perspectives, and views of the task (Jehn, 1995). Task conflict is of particular interest because, unlike most types of conflict, it has been argued to be conducive to team performance (Amason, 1996). Meta-analyses, however, indicate that task conflict is weakly related to team performance, on average (e.g., O’Neill, Allen, & Hastings, 2013).

A potential limitation of task conflict research is that it ignores whether resolutions of those conflicts were actually reached. This is an important issue, because a team with unresolved conflicts may experience impasses, deadlocks, and barriers to forward movement, whereas a team with resolved conflicts likely integrates, combines, and synthesizes divergent views and achieves consensus on how best to proceed (see Robey, Farrow, & Franz, 1989). Our review of the literature presented below, however, suggests that empirical research examining both antecedents and consequences of task conflict resolution is lacking. In light of the potential importance of task conflict resolution, we consider both innovation and task performance as possible team outcomes in a sample of engineering design project teams. We also investigate potential antecedents (i.e., input variables)

---

Thomas A. O’Neill, Department of Psychology, University of Calgary, Calgary, Alberta, Canada; Natalie J. Allen, Department of Psychology, Western University, London, Ontario, Canada.

Correspondence concerning this article should be addressed to Thomas A. O’Neill, Department of Psychology, University of Calgary, 2500 University Drive, N.W., Calgary, AB Canada T2N 1N4, P: 403-220-5207. E-mail: toneill@ucalgary.ca

that could be useful for pinpointing the conditions in which task conflicts are most likely to be resolved. Specifically, in this research, we focus on team-level Manipulativeness, Narcissism, and Secondary Psychopathy. These traits are similar, although not identical, to the “Dark Triad” (Paulhus & Williams, 2002), which includes a common core of selfishness, remorselessness, disagreeableness, and aggressiveness (Jonason, Webster, Schmitt, Li, & Crysel, 2012). Although Judge and LePine (2007) called for team research on dark traits, current literature has primarily focused on the Five-Factor Model (Bell, 2007; Driskell, Goodwin, Salas, & O’Shea, 2006).

Given our examination of team inputs (personality composition), processes (task conflict resolution), and outcomes (innovation and task performance), the current research design presents the opportunity to investigate the role of task conflict resolution as a process variable that mediates the effects of team personality composition on team outcomes. As LePine, Buckman, Crawford, and Methot (2011) noted in their recent review, the intervening processes and mechanisms linking team personality inputs to team outcomes are understudied and largely unknown. This study aims to address that gap.

## Theoretical Development

### Task Conflict Resolution

In the current research we draw attention to the issue of task conflict resolution and investigate its associations with team inputs and outcomes. As described by Robey et al. (1989), “conflict resolution is apparent when members have fewer objections and indicate more agreement with each other . . . it indicates a solution that is acceptable to the entire group” (p. 1174). Wall and Callister (1995) noted that conflict resolutions “may be in the form of explicit or tacit agreement” (p. 525), and Kuhn and Poole (2000) viewed conflict resolution as “coming to a mutually agreeable solution” (p. 570). We define task conflict resolution as *the extent to which different opinions, viewpoints, and perspectives were resolved*.

As mentioned, recent meta-analyses summarizing the substantial body of research involving task conflict and team performance report small effect sizes (e.g., de Wit et al., 2012; O’Neill et

al., 2013). *Resolving* task conflicts, on the other hand, could be critical to team outcomes because team members have encountered and adequately addressed divergent viewpoints regarding the direction in which the team ought to proceed. Compared with research on conflict types, there is scant consideration of the extent to which a team actually resolves conflicts, the type of conflicts resolved, and whether resolutions predict team outcomes (e.g., De Dreu, 2011). We also note that there is research on the strategies, processes, and styles of negotiating conflicts, resolving disputes, and allocating scarce resources (e.g., Behfar, Peterson, Mannix, & Trochim, 2008; Katz & Kahn, 1978; Rubin, Pruitt, & Kim, 1994), but there is little empirical research informing our understanding of the antecedents and consequences of task conflict resolution.

Theoretical work on conflict resolution by Robey and colleagues (i.e., Robey & Farrow, 1982; Robey et al., 1989) deals with factors leading to satisfactory conflict resolution (see also Barki & Hartwick, 1994). Notable findings from that research suggest participation leads to power and influence, which in turn is positively associated with satisfactory resolution of conflict. However, those studies were conducted at the individual level and, therefore, they ignore whether the team, as a collective, was satisfied with resolutions. Furthermore, the studies found the different types of conflict (e.g., task, relationship), and conflict resolution is treated as the final outcome rather than as a process affecting team innovation and task performance. In a notable exception, Jehn, Greer, Levine, and Szulanski (2008) offer a three-item measure of task conflict resolution *efficacy*, which was related positively to team viability and emergent states. However, the construct pertains to “the belief that conflict can be easily resolved” (p. 465) rather than actual conflict resolutions. Greer, Jehn, and Mannix (2008) used single-item scales to measure each of task, relationship (i.e., interpersonal incompatibilities), and process (e.g., work-structuring incompatibilities) conflict resolutions. Arguably, multi-item scales might be more reliable and could provide better coverage of the content domains, and consideration of team outcomes (e.g., team task performance) might shed new light on the importance of task conflict resolution.

## Task Conflict Resolution and Team Outcomes

Team innovation involves developing a creative or novel solution that is also practical and feasible in its development and application (e.g., Amabile, 1983; Drach-Zahavy & Somech, 2001a, 2001b), whereas team task performance refers to the overall quality of the team's task-related output (e.g., Hackman, 1987). This distinction suited the context of the current study. Specifically, the objective of the engineering project teams involved was to develop a novel and effective approach to demonstrating how an environmentally friendly technology creates renewable energy (e.g., team innovation), and to deliver an extensive report documenting the team's design and the prototype's capabilities (team task performance).

Teams capable of resolving task conflicts are likely to perform well and reach high levels of innovation. Drawing from social interdependence theory, compatible team-member goals appear to facilitate effective conflict handling, which in turn results in identifying novel and high-quality solutions (see Deutsch, 2006). Additionally, based on information processing theory, we suggest unresolved conflicts may arouse the autonomic nervous system, thereby inhibiting critical analysis of the task and creative thinking (e.g., Carnevale & Probst, 1998). Indeed, teams with positive interpersonal relationships exhibit greater creativity than do teams in conflict (e.g., Craig & Kelly, 1999).

In project teams of the nature studied here, we anticipated this would be particularly important because meeting milestones and keeping forward momentum was essential to team innovation and team task performance (see Chiocchio, 2007). Behfar et al.'s (2008) qualitative study of project teams found that high-performing teams used pluralistic strategies involving either a preemptive focus on addressing potential conflicts, or quickly and efficiently containing conflicts by invoking rules to resolve conflicts (e.g., majority rules). Teams failing to address and to resolve different viewpoints are likely to experience stalls or deadlocks that inhibit smooth progress. In sum, we predict the following:

H<sub>1a</sub>: Task conflict resolution will be positively correlated with team innovation.

H<sub>1b</sub>: Task conflict resolution will be positively correlated with team task performance.

## Input-Process-Output (IPO) Model and Team Personality Composition

In the current research we applied the well-known IPO model of team effectiveness (see McGrath, 1984). The IPO model is best considered a heuristic that organizes theorizing about how team inputs (e.g., personality) affect team processes (e.g., task conflict resolution) and, in turn, team outcomes (e.g., innovation, task performance). In the context of the present research, task conflict resolution is theorized to be a mechanism by which personality composition impacts team outcomes.

Variables organized under the IPO framework are conceptualized at the team level and, therefore, personality variables must be aggregated prior to hypothesis testing. Operationalizing traits using the within-team mean, as we do here, invokes an additive theory of aggregation (see Chan, 1998). Based on the additive theory, the arithmetic mean of team member scores is most relevant because it is the overall "abundance" of the trait, within the team, that is expected to be most consequential to the team's work (e.g., LePine, Hollenbeck, Ilgen, & Hedlund, 1997; LePine, 2003; O'Neill & Allen, 2011).

## Dark Traits as Antecedents of Task Conflict Resolution

One potential approach is the "Dark Triad," which encompasses Narcissism, Machiavellianism, and Psychopathy (Paulhus & Williams, 2002). Narcissism describes people who hold feelings of superiority, grandiosity, and entitlement (e.g., Ames, Rose, & Anderson, 2006; Raskin & Terry, 1988). Rather than focusing on Machiavellianism, we chose a central dimension of Machiavellianism referred to as Manipulativeness (see Paunonen & Jackson, 2000). Manipulativeness involves the willingness to use of flattery, ingratiation, and deception to advance personal agendas (see Paunonen, Haddock, Forsterling, & Keinonen, 2003). We view the content of the Manipulativeness dimension of Machiavellianism as highly relevant to the teams' capacities to resolve conflicts. Although

individuals high on Manipulativeness use impression management techniques to advance their own agenda, they do not necessarily possess the charm, appeal, and magnetism of Machiavellians (cf. Paunonen & Jackson, 2000). Finally, we considered Secondary Psychopathy, which involves impulsivity, lack of goal commitment, delinquency, irresponsibility, and difficulties establishing relationships (e.g., Levenson, Kiehl, & Fitzpatrick, 1995; McHoskey, Worzel, & Szyarto, 1998). Although the current traits do not overlap perfectly with the Dark Triad, the Dark Triad literature is relevant to formulating our hypotheses.

Affective events theory (AET; Weiss & Cropanzano, 1996) is a framework for understanding how Manipulativeness, Narcissism, and Secondary Psychopathy may interfere with the resolution of task conflict. One proposition that is consistent with AET is that personality affects the occurrence of events at work and how those events translate into affective experiences, which in turn have implications for behavior (see Wegge, van Dick, Fisher, West, & Dawson, 2006). Common tendencies of individuals high on dark traits similar to those considered in this study include the use of adversarial and disruptive behaviors, and engagement in malevolent interpersonal dealings (e.g., Jonason & Webster, 2012). As such, these individuals would likely exacerbate the challenge of effectively resolving task conflicts by arousing negative emotions. This could occur through their use of "hard tactics" such as threats, punishments, and manipulations (e.g., Jonason, Slomski, & Partyka, 2012), lack of self-control and understanding of behavioral consequences (e.g., Jonason & Tost, 2010), emotional intelligence deficits (e.g., Jordan & Troth, 2004; Petrides, Vernon, Schermer, & Veselka, 2011), and empathy limitations (e.g., Jonason & Krause, 2013; Wai & Tiliopoulos, 2012). Negative emotions and arousal in teams high on dark traits would interfere with effective information processing (e.g., Carnevale & Probst, 1998), which is likely to detract from problem-focused discussions aimed at resolving task conflicts. Accordingly, we offer the following predictions:

H<sub>2</sub>: Mean Manipulativeness, Narcissism, and Secondary Psychopathy will be negatively correlated with task conflict resolution.

In the current study we measured dark traits, conflict, and outcomes at three distinct

and sequential time periods. Although we do not use an experimental design and, as a consequence, causality cannot be unequivocally determined, the ordering of measurement coincides with the likely temporal ordering of these variables. Specifically, we theorize that personality would most likely be an antecedent (not a mediator or outcome), and that task conflict resolution would most likely precede team outcomes.

H<sub>3</sub>: Task conflict resolution will mediate the relationships between team personality variables (Manipulativeness, Narcissism, and Secondary Psychopathy) and team outcome variables (innovation and performance).

### Dark Traits and Team Outcomes

Although we argued for an indirect linkage involving the dark traits and team outcomes, theory also supports direct effects. Manipulative individuals engage in impression management strategies involving ingratiation and self-promotion (see Paunonen et al., 2003), which, when viewed as insincere, tend to be associated with distrust and skepticism (e.g., Baron, 1986). As several team members attempt to manipulate each other, the team atmosphere would likely be rife with suspicion, which would not likely promote team collaboration, innovation, or performance. Team members high on Narcissism might have excessively optimistic views about the quality of their work or their own contributions even in light of evidence to the contrary. Furthermore, narcissistic individuals are seen by their group members as domineering, ingenuous, and lacking in interpersonal intelligence (e.g., Rauthmann, 2012), and a team of such individuals would not likely be effective. Finally, teams with members high on Secondary Psychopathy would likely derail the team's progress because of their members' impulsivity, aggressiveness, and emotional instability (cf. McHoskey et al., 1998; Paulhus & Williams, 2002). In sum, we propose the following:

H<sub>4a</sub>: Mean Manipulativeness, Narcissism, and Secondary Psychopathy will be negatively correlated with team innovation.

H<sub>4b</sub>: Mean Manipulativeness, Narcissism, and Secondary Psychopathy will be negatively correlated with team task performance.

## Method

### Participants and Teamwork Context

Data were collected from 344 team members working in 81 engineering project design teams for 6.5 months. The participants were students in an engineering design course at a large Canadian university. Teams had four ( $n = 61$ ) or five ( $n = 20$ ) members, and 85% of participants were male ( $M$  age = 18.6 years,  $SD = 2.4$ ). Teams were formed pseudorandomly. Specifically, individuals were asked to rank order their proficiency on “skill sets” (i.e., communication, computer, planning, hands on), and then teams were built such that they had one individual with a ranking of “1” on each dimension. The course instructors preferred this method to full random assignment.

Our observations indicate teams took great pride in designing their devices and presenting them at a high-profile design showcase in which the city mayor, university president, engineering dean and faculty, and the public were in attendance. This “major design project” was the focus of team innovation and team task performance in the current research. The projects were complicated and inherently high on knowledge interdependence because team members tended to become experts in particular design areas critical to the prototype’s innovation, functionality, and documentation. Thus, the teams in this sample fit well within the accepted definition of work teams: highly interdependent members working on consequential tasks and having a collective purpose (e.g., Allen & West, 2005; Hackman, 1987).

### Procedure and Timeline

Participants were invited to complete personality and conflict surveys during weekly design classes. Personality data were collected in the first design class before participants were assigned to their teams; conflict processes were assessed 6 months later, 2 weeks before completing the major design project. Team innovation ratings were based on prototype photos and written descriptions collected a week before final project reports were due. Team task performance was assessed after the major design project was submitted at the 6.5-month mark.

## Measures

**Time 1: Personality.** Manipulativeness was measured using Paunonen’s (2002) scale from the Supernumerary Personality Inventory (SPI). The scale contains 15 items with 5-point Likert-type response options ( $\alpha = .70$ ). Example items are as follows: *By flattering someone I can make him or her more apt to agree with me* and *I will pretend to be extremely interested in what a person is saying in order to get something*. Narcissism was measured using the 16-item forced-choice form of the Narcissistic Personality Inventory ( $\alpha = .67$ ; Ames et al., 2006). Specifically, respondents were asked to choose which of two statements are most like them, such as *I really like to be the center of attention* and *It makes me uncomfortable to be the center of attention*. Another item set was *People sometimes believe what I tell them* and *I can make anybody believe anything I want them to*. Secondary Psychopathy was measured using 10 items from the Levenson Self-Report Psychopathy Scale ( $\alpha = .63$ ; Levenson et al., 1995) with 5-point Likert-type response options. Example items are *I find myself in the same kinds of trouble, time after time* and *I have been in a lot of shouting matches with other people*.

**Time 2: Task conflict resolution and task conflict.** To measure task conflict resolution, we observed teams in their engineering design laboratories over a 6 month period. Extensive diary notes were taken during the approximate 100 hours we spent, overall, observing various teams receiving instructions, training, and feedback, and during brainstorming, decision-making exercises, and design work. Specifically, the first author attended 4 of the 8 weekly design labs on a rotating basis during which teams were engaged with various group activities including the focal project. Through this informal team task analysis, we identified four key project milestones: a) problem definition, b) deciding on the team’s design concept, c) developing the team’s prototype, and d) preparing the team presentation. Next, we adopted the following general form across four items: *the extent to which different opinions, viewpoints, and perspectives were resolved before settling on the team’s \_\_\_\_\_*. For each item, the blank contained *problem definition, design concept, prototype, and presentation* ( $\alpha = .89$ ), respectively. Jehn’s (1995) four-item measure



of task conflict was also administered ( $\alpha = .89$ ). Scale anchors for both measures ranged from 1 (*A very small amount*) to 5 (*A lot*).

**Time 3: Team innovation.** Team innovation was assessed using Amabile's (1983) consensual assessment technique, which involves collecting innovation judgments from domain-relevant experts. Five expert judges who were engineering graduate students rated the innovation of each design prototype. The targets of experts' ratings were short, written descriptions and  $4 \times 6$  in. color photographs of the team prototypes. The relative percentile method (RPM) formed the basis of the rating scale employed to assess team innovation (see Goffin, Gellatly, Paunonen, Jackson, & Meyer, 1996). The current application of the RPM required judges to rate each team's level of innovation on a scale ranging from zero to 100. Team innovation was scored as the average of each judge's rating of each team prototype. An intraclass correlation (ICC) was computed for assessing the reliability of multiple raters rating team innovation on a continuous scale [ICC(2) = .71; Shrout & Fleiss, 1979]. We note that the experts were course teaching assistants leading different design studios, and they were each familiar with a unique set of approximately 10 of the 81 projects. However, all judges rated all projects in a fully crossed design. Therefore, any possible bias introduced by familiarity with the projects would likely be limited after averaging across all five judge ratings, and the strong interrater reliability supports this view.

**Time 4: Team task performance.** Teams were required to submit reports of approximately 100 pages containing background research, documentation of the theoretical and observed capabilities of the team's design con-

cept, CAD sketches, mathematical modeling, and implications. The assessment of team task performance consisted of ratings provided by experienced course instructors. Because raters did not rate the same teams, we controlled for the possibility of raters using different performance distributions by  $z$ -standardizing within raters (which is common practice in similar situations; O'Neill & Allen, 2012; Wageman & Gordon, 2005).

## Results

### Aggregation

Task conflict resolution and task conflict are "shared-unit" constructs; therefore, within-team agreement and between-team heterogeneity is required to support construct validity prior to aggregation (Kozlowski & Klein, 2000). According to Bliese (2000), ICCs indicate the proportion of variance explained by group membership [that is, ICC(1)] and the reliability of the group mean [that is, ICC(2)]. Table 1 contains ICCs for the shared-unit constructs of task conflict and task conflict resolution, which were all significantly greater than zero. Furthermore, Jehn et al. (2008) interpreted ICCs of this magnitude (and lower) as sufficient for justifying aggregation. The personality variables were aggregated using the mean but were not expected to demonstrate within-group agreement because teams were formed pseudorandomly, traits are configural unit properties (i.e., not shared; Kozlowski & Klein, 2000), and their function is theorized to be additive across team members (LePine, 2003; O'Neill & Allen, 2011).

Table 1  
Means, Standard Deviations, ICCs, and Correlations Involving Team-Level Study Variables

Variable	ICC(1)	ICC(2)	<i>M</i>	<i>SD</i>	1	2	3	4	5	6
1. Manipulativeness			2.91	0.21						
2. Narcissism			4.95	1.38	.40*					
3. Secondary psychopathy			2.34	0.23	.17	.29*				
4. Task conflict	.24*	.54*	2.13	0.51	.17	.21	.26*			
5. Task conflict resolution	.14*	.40*	3.91	0.55	-.02	-.12	-.30*	-.39*		
6. Team innovation			53.96	17.38	.05	-.06	-.21	-.11	.14	
7. Team task performance			0.00	0.96	-.06	.08	-.30*	-.18	.30*	.37*

Note.  $n = 81$ .

\*  $p < .05$ .

## Tests of Hypotheses

We predicted task conflict resolution would be positively related to team innovation ( $H_{1a}$ ). Although in the predicted direction, this relation did not reach significance (see Table 1). Task conflict resolution was expected to positively correlate with team task performance ( $H_{1b}$ ). As this was the case,  $H_{1b}$  was supported.

We anticipated that Manipulativeness, Narcissism, and Secondary Psychopathy would be negatively related to task conflict resolution ( $H_2$ ). Although mean Manipulativeness and Narcissism were not related to task conflict resolution, there was a significant negative relation involving Secondary Psychopathy (see Table 1). We also expected that the effect of the dark personality traits on team outcomes would be mediated by task conflict resolution ( $H_3$ ). Preacher and Hayes (2004) recommended the bootstrapping approach for testing the significance of the indirect effect of the predictor on the outcome through the intervening variable. Specifically, bias-corrected upper and lower 95% confidence intervals were calculated using 1000 bootstrapped samples. Analyses supported the significance of the indirect effect for one model (see Figure 1): Secondary Psychopathy  $\rightarrow$  task conflict resolution  $\rightarrow$  team task performance (unstandardized indirect effect =  $-.28$ ,  $CI_{95\%} = -.68$  to  $-.06$ ). This indicates a one-unit increase in mean Secondary Psychopathy was related to a .28 decrease in team task performance through task conflict resolution. Furthermore, the direct effect of Secondary Psychopathy on team task performance remained significant after taking into account the indirect effects, suggesting Secondary Psychopathy has both direct and indirect paths to team task performance.<sup>1</sup>

Regarding team outcomes, Manipulativeness, Narcissism, and Secondary Psychopathy should be related negatively to team innovation ( $H_{4a}$ ). However, none of the relations reached conventional levels of significance (see Table 1). Turning to team task performance ( $H_{4b}$ ), of the three traits examined, a significant negative association was found only for Secondary Psychopathy.

## Supplemental Analyses

Given the pattern of results involving personality, Secondary Psychopathy appeared to be the most important predictor of the teamwork

variables (see Table 1). One procedure for comparing the importance of predictors is relative importance analysis (e.g., Johnson, 2000), which overcomes the limitations of both zero-order correlations as well as betas from standard multiple regression (e.g., Johnson & LeBreton, 2004). Table 2 contains the results of the relative importance analysis as well as overall model  $R^2$  values. Model  $R^2$  values indicate that the personality traits, as a set, accounted for 12% of the variance in team task performance and 10% of the variance in task conflict resolution. Overall prediction of task conflict and team innovation was nonsignificant. In terms of the relative importance of the three personality traits, the results clearly indicate that Secondary Psychopathy was dominant. Of the proportion of variance explained by the personality variables as a set, Secondary Psychopathy accounted for 57% (task conflict), 78% (team task performance), 82% (team innovation), and 91% (task conflict resolution). Thus, Secondary Psychopathy was the most important predictor of the team process and outcome variables studied.

## Discussion

The overarching purpose of this study was to investigate the role of task conflict resolution as a team process linking dark personality traits to team outcomes. This is notable because the actual resolution of task conflict has received much less consideration than has the occurrence of task conflict, yet resolution could be at least as important. Taken together, the set of hypotheses we forwarded involving these relations received partial support. In our view, the contribution of the current study is threefold. First, compared with task conflict, findings for the resolution of task conflict appear more intriguing and theoretically informative. Task conflict resolution was significantly related to team task performance, and it was a mediator between Secondary Psychopathy and team task performance. This is novel given the dearth of past

<sup>1</sup> For completeness, we considered task conflict as a mediator and as a moderator in moderated mediation analyses. The moderated mediation analyses examined the role of task conflict moderating either the effect of task conflict resolution on team outcomes or personality traits on task conflict resolution (Hayes, 2013). We did not find any significant effects for the function of task conflict in any of these roles.

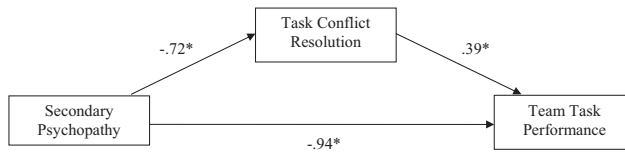


Figure 1. Mediation model with unstandardized path coefficients. All paths in the model are significant including the direct path from Psychopathy to team task performance after controlling for the indirect effect through task conflict resolution. The indirect effect was  $-.28$ . \*  $p < .05$ .

research on task conflict resolution and the existing focus on task conflict levels. Indeed, the authors of recent meta-analyses found 83 (O'Neill et al., 2013) and 116 (de Wit et al., 2012) articles investigating conflict types and team performance. Consistent with information processing theory, results from those meta-analyses suggested, with few exceptions, that task conflict is either unrelated to or harmful for team performance (as it was in the present research). Accordingly, a continued focus on task conflict itself may be limiting (see also DeChurch, Mesmer-Magnus, & Doty, 2013). Our findings are important for theory and practice because they imply that the resolution of task conflict may be a critical factor. This fits well within social interdependence theory, as interdependence often leads to conflict and, therefore, the manner in which the conflict is managed may be most important (see Johnson,

2003; Tjosvold, 1998). It is also consistent with the project management literature (e.g., Chiocchio & Lafrenière, 2009) given that resolved task conflict is needed to progress smoothly and efficiently through project stages. Thus, it may be advantageous for researchers and practitioners to emphasize strategies that help teams work through task conflicts rather than to encourage teams to engage in task conflict without addressing its resolution.

A second novel contribution of the current research involves the examination of team personality composition from the perspective of dark traits (cf. Judge & LePine, 2007). Relative to Manipulativeness and Narcissism, Secondary Psychopathy dominated the prediction of task conflict resolution, team innovation, and team task performance. Interestingly, Secondary Psychopathy exhibited stronger relations with team task performance than did any of the Big Five,

Table 2  
Relative Importance Analysis of Team-Level Personality Variables

Criterion and statistic	Manipulativeness	Narcissism	Secondary psychopathy	R <sup>2</sup>
Task conflict				
β	.08	.12	.21 <sup>†</sup>	.09
Raw relative weight	.02	.03	.05	
Regression weight as a percentage of R <sup>2</sup>	16	28	57	
Task conflict resolution				
β	.05	-.05	-.30*	.10*
Raw relative weight	.00	.01	.09	
Regression weight as a percentage of R <sup>2</sup>	1	8	91	
Team innovation				
β	.11	-.04	-.22 <sup>†</sup>	.06
Raw relative weight	.01	.00	.05	
Regression weight as a percentage of R <sup>2</sup>	11	7	82	
Team task performance				
β	-.09	.21	-.32*	.12*
Raw relative weight	.01	.02	.10	
Regression weight as a percentage of R <sup>2</sup>	5	17	78	

Note.  $n = 81$ .  
\*  $p < .05$ .

in terms of uncorrected correlations, reported in past meta-analyses (e.g., Bell, 2007). Characteristic patterns of behavior of individuals high on Secondary Psychopathy include difficulties relating to others, emotional outbursts, impulsivity, and lack of goal commitment (Levenson et al., 1995; McHoskey et al., 1998). Consistent with theories such as affective events theory, teams of individuals with high Secondary Psychopathy likely experienced frequent and repeated cycles of negative affective interactions, potentially resulting in reinforcing downward spirals (cf. Peterson & Behfar, 2003). For example, individuals high on Secondary Psychopathy often use social influence tactics involving coercion, debasement, and threats (Jonason et al., 2012); tend to be dishonest and lack humility (Lee & Ashton, 2005); exhibit self-control difficulties (Jones & Paulhus, 2011; Jonason & Tost, 2010); appear to engage in various forms of aggression, cheating, and bullying (Boddy, 2010; Coyne & Thomas, 2008); and tend to hold social dominance beliefs (Hodson, Hogg, & MacInnis, 2009).

Given the above, placing a number of individuals high on Secondary Psychopathy in a team will likely lead to numerous negative affective interactions, disputes, and arguments. It may be advisable, therefore, to build teams such that overall levels of Secondary Psychopathy are as low as possible. One way to accomplish this involves the “seeding” method outlined by Humphrey, Hollenbeck, Meyer, and Ilgen (2007). When a trait is known to combine *additively* (such that increasing or decreasing the team mean on that trait is helpful for performance), they would suggest assessing the trait and distributing the highest scorers across teams rather than concentrating them in any one team. With criterion validity evidence and an analysis of the personality-based work requirements suggesting the trait is job relevant (see Arthur, Edwards, Bell, Villado, & Bennett, 2005), the case for the legal defensibility of using the test to make personnel decisions with seeding methods could be strengthened (cf. Society for Industrial-Organizational Psychology *Principles*, 2003, and local employment law).

A third contribution is that we identify task conflict resolution as a mechanism linking team personality composition, specifically mean levels of Secondary Psychopathy, to team task performance. This suggests a team of individu-

als high on Secondary Psychopathy will tend to experience failures in resolving task conflicts, which, in turn, results in difficulties moving the project forward and achieving high task performance. Thus, although meta-analyses highlight the association between personality and performance in teams (e.g., Bell, 2007), in the current study we report on one of the few empirical findings supporting a linking mechanism through which team personality composition influences team performance. As mentioned above, teams high on mean Secondary Psychopathy contain individuals who are likely hostile, aggressive, impulsive, and unable to remain committed to goals. It appears teams of such individuals experience barriers to resolving task conflicts and to finding agreement regarding avenues toward high-quality task work. In turn, this was associated with reductions in team effectiveness. Importantly, both indirect and direct connections to team task performance were supported in the model, suggesting mean Secondary Psychopathy has at least two distinct paths through which it impacts team task performance. Investigating additional team processes, such as psychological safety (Edmondson, 1999), communication quality and quantity (Barrick, Stewart, Neubert, & Mount, 1998), and other potentially relevant processes are promising avenues for future research.

### On the Role of Mixed Motives

It is worth considering whether a lack of mixed motives and hierarchical goal structures might explain why Manipulativeness and Narcissism were unrelated to team conflict processes and team outcomes. The underlying rationale for the relevance of Manipulativeness and Narcissism is that they should be related to individuals advancing their own objectives, agendas, and goals at the expense of pursuing the team’s collective purpose. Findings from studies examining social interdependence theory indicate that this situation leads to intrateam competition, thereby contributing to selective information sharing, an argumentative approach to differences of opinion, distrust, and a closed-minded interpretation of new information (see reviews by Johnson, 2003; Johnson & Johnson, 1989). But whether team members had proself motives is somewhat of an open question.

Our observations and familiarity with the teamwork context indicate that members quite likely experienced varying degrees of divergence in individual- and team-level motives. First, participants had not yet specialized in a particular subfield of engineering, and therefore the team was faced with identifying one or two subfields on which their prototype would be based (e.g., mechanical, electrical, software, urban, robotics). Given that team members often had different interests, this decision likely created some variance between individual and team preferences. In turn, we expect that this creates the conditions for manipulateness, conceit, and selfishness to inhibit conflict resolution and joint effort toward a collective purpose. Second, team members likely had divergence on the level of quality and detail they felt was needed, how many hours and meetings to invest, the nature of member roles and timelines, and so forth. Here again, manipulative and egotistical behaviors would hinder integration, compromise, and consensus, thereby contributing to conflict and poor team outcomes. To the extent that team members had different preferences and goals such as those described here, it is almost certain that some individuals' objectives would be at odds with team objectives. As a result, we speculated that teams high on Manipulateness or Narcissism would suffer. The extent to which individual and group motives and goals were distinct, however, is difficult to pinpoint, and future studies investigating Manipulateness, Narcissism, and related traits in teams may benefit from considering goal hierarchies and goal incongruences (e.g., De-Shon, Kozlowski, Schmidt, Milner, & Wiechmann, 2004).

### Future Research

In light of the current study, we view team task conflict resolution as a promising avenue for advancing team conflict research. In addition to the positive association with team task performance, team task conflict resolution exhibited a negative relation with task conflict. This suggests a number of teams experienced difficulties with persistent and unresolved task conflicts, which may well have reduced their ability to function effectively. Correlations between task conflict resolution and the dark personality traits were in the negative direction,

which supports the construct validity of the task conflict resolution measure. Thus, future research on task conflict resolution would seem fruitful.

A second direction for future research involves studying *how* conflict is resolved. There is a voluminous literature on conflict management approaches (e.g., Deutsch, 1973) and styles (e.g., Rahim, 1983), providing evidence that cooperative goals and integrative styles are associated with resolved conflicts (see Lewicki, Weiss, & Lewin, 1992). But there is a need for research on how to promote effective approaches to conflict management and resolution, and what specific strategies and behaviors lead to effective resolutions. For example, would it matter whether disagreements are resolved through voting, autocracy, or avoidance? Recently, Behfar et al. (2008) identified specific strategies used by teams to resolve conflicts. Characteristic of high-performing and highly satisfied teams was equity-focused resolution (e.g., work assignments based on skill). Characteristic of low-performing and unsatisfied teams was adhocracy-focused resolution (e.g., avoidance of conflict and group members in general). Tactics such as voting and majority rules were associated with high performance but, interestingly, low satisfaction. On the other hand, equality rules that promoted cohesiveness were associated with low performance and high satisfaction. What might drive this research forward is scale development and empirical examination of how these different resolution strategies impact resolution itself, satisfaction with the resolution, and team outcomes. This could deepen our understanding of the specific behaviors that are needed for effective conflict resolution.

In the way of a third avenue for future research, it may be valuable to consider relationship conflict resolution and process conflict resolution. The current research focused on task conflict resolution as a first step, as this type of conflict has been studied extensively and has been described as the "productive" form of conflict (e.g., De Dreu, 1997). It seemed to us that task conflict resolution may be promising, and our research provides cause for optimism in this regard. Greer et al. (2008) found process conflict resolution moderated the relation between early process conflict and later process, task, and relationship conflict such that resolutions

buffered continuing conflict occurrences. These findings are encouraging despite being based on single-item conflict resolution scales; therefore, development of multiitem scales would likely be advantageous.

### Limitations and Conclusion

First, although the dark traits investigated in the current research are similar to the Dark Triad, and we draw from the Dark Triad literature, there are some important differences. We focused on Manipulativeness, which is a feature of Machiavellianism but is not identical. Machiavellianism is a broader trait including charm, charismatic tendencies (see Christie, Geis, & Berge, 1970), and self-control (see Jones & Paulhus, 2011), and the fact that Manipulativeness was not predictive of teamwork variables does not rule out the potential predictiveness of Machiavellianism. We also considered Secondary Psychopathy because we expected that it would have little overlap with Manipulativeness relative to Primary Psychopathy. However, the Dark Triad contains coverage of Primary Psychopathy, such as cynicism and callousness (Jonason & Webster, 2010), which is strongly correlated with Machiavellianism (e.g., McHoskey et al., 1998). Therefore, we had a theoretical interest in considering traits that are slightly different from the Dark Triad, and our findings should not be assumed to fully capture the effects of the Dark Triad on team conflict processes or team outcomes. Second, despite the use of a cross-sectional design with three measurement periods that corresponded to our theoretical causal ordering of the relevant variables, the mediation analyses do not allow definitive “cause-and-effect” conclusions (Judd & Kenny, 2010). Thus, experiments with full randomization to conditions should be conducted. Third, replication of the current research is necessary to shed light on the robustness and generalizability of the findings. Fourth, although we used a unidimensional measure of Narcissism, this trait consists of subdimensions concerning entitlement, leadership, self-absorption, and superiority (Emmons, 1987). These subdimensions may have differential relations with teamwork variables, and therefore the current study may not have detected nontrivial facet-level findings (cf. O’Neill & Allen, 2011). On the other hand, Manipulativeness is a unidi-

mensional construct (Paunonen & Jackson, 2000) and Secondary Psychopathy is a narrower measure of a broad Psychopathy construct (Ross, Bye, Wrobel, & Horton, 2008), suggesting the issue may not be as relevant for these latter traits.

The current research emphasizes the role of task conflict resolution and provides evidence of a positive relation with team task performance. Furthermore, task conflict resolution was found to operate as an indirect linking mechanism between mean Secondary Psychopathy and team task performance. Thus, it seems new research focusing on the resolution of conflict, rather than its levels, could be useful for enriching our understanding of the role of conflict processes in organizational work teams. Finally, in addition to influencing both task conflict resolution and overall task conflict levels, mean Secondary Psychopathy exhibited negative relations with team task performance, and at a marginally significant level, team innovation. Thus, further investigations of Secondary Psychopathy may be of considerable value for advancing theory on team personality composition.

### References

- Allen, N. J., & West, M. A. (2005). Selecting for teamwork. In A. Evers, N. Anderson, & O. Voskuil (Eds.), *The Blackwell handbook of personnel selection* (pp. 476–494). Oxford, UK: Blackwell Press.
- Amabile, T. M. (1983). *The social psychology of creativity*. New York, NY: Springer-Verlag. doi:10.1007/978-1-4612-5533-8
- Amason, A. C. (1996). Distinguishing the effects of functional and dysfunctional conflict on strategic decision making: Resolving a paradox for top management teams. *Academy of Management Journal*, 39, 123–148. doi:10.2307/256633
- Ames, D. R., Rose, P., & Anderson, C. P. (2006). The NPI-16 as a short measure of narcissism. *Journal of Research in Personality*, 40, 440–450. doi:10.1016/j.jrp.2005.03.002
- Arthur, W., Edwards, B. D., Bell, S. T., Villado, A. J., & Bennett, W. (2005). Team task analysis: Identifying tasks and jobs that are team based. *Human Factors*, 47, 654–669. doi:10.1518/001872005774860087
- Barki, H., & Hartwick, J. (1994). User participation, conflict, and conflict resolution: The mediating roles of influence. *Information Systems Research*, 5, 422–438. doi:10.1287/isre.5.4.422

- Baron, R. A. (1986). Self Presentation in Job Interviews: When there can be "too much of a good thing". *Journal of Applied Social Psychology, 16*, 16–28. doi:10.1111/j.1559-1816.1986.tb02275.x
- Barrick, M. R., Stewart, G. L., Neubert, M. J., & Mount, M. K. (1998). Relating member ability and personality to work-team processes and team effectiveness. *Journal of Applied Psychology, 83*, 377–391. doi:10.1037/0021-9010.83.3.377
- Behfar, K. J., Perterson, R. S., Mannix, E. A., & Trochim, M. K. (2008). The critical role of conflict resolution in teams: A close look at the links between conflict type, conflict management strategies, and team outcomes. *Journal of Applied Psychology, 93*, 170–188. doi:10.1037/0021-9010.93.1.170
- Bell, S. T. (2007). Deep-level composition variables as predictors of team performance: A metaanalysis. *Journal of Applied Psychology, 92*, 595–615. doi:10.1037/0021-9010.92.3.595
- Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 349–381). San Francisco, CA: Jossey-Bass.
- Boddy, C. R. (2010). Corporate psychopaths, bullying and unfair supervision in the workplace. *Journal of Business Ethics, 100*, 367–379. doi:10.1007/s10551-010-0689-5
- Carnevale, P. J., & Probst, T. M. (1998). Social values and social conflict in creative problem solving and categorization. *Journal of Personality and Social Psychology, 74*, 1300–1309. doi:10.1037/0022-3514.74.5.1300
- Chan, D. (1998). Functional relations among constructs in the same content domain at different levels of analysis: A typology of composition models. *Journal of Applied Psychology, 83*, 234–246. doi:10.1037/0021-9010.83.2.234
- Chiocchio, F. (2007). Project team performance: A study of electronic task and coordination communication. *Project Management Journal, 38*, 97–109.
- Chiocchio, F., & Lafrenière, A. (2009). A project management perspective on student's declarative commitments to goals established within asynchronous communication. *Journal of Computer Assisted Learning, 25*, 294–305. doi:10.1111/j.1365-2729.2009.00310.x
- Christie, R., Geis, F. L., & Berger, D. (1970). *Studies in Machiavellianism* (p. ). New York, NY: Academic Press.
- Coyne, S. M., & Thomas, T. J. (2008). Psychopathy, aggression, and cheating behavior: A test of the cheater-hawk hypothesis. *Personality and Individual Differences, 44*, 1105–1115. doi:10.1016/j.paid.2007.11.002
- Craig, T. Y., & Kelly, J. R. (1999). Group cohesiveness and creative performance. *Group Dynamics: Theory, Research, and Practice, 3*, 243–256. doi:10.1037/1089-2699.3.4.243
- DeChurch, L. A., Mesmer-Magnus, J. R., & Doty, D. (2013). Moving beyond relationship and task conflict: Toward a process-state perspective. *Journal of Applied Psychology, 98*, 559–578. doi:10.1037/a0032896
- De Dreu, C. K. W. (1997). Productive conflict: The importance of conflict management and conflict issue. In C. K. W. De Dreu & E. Van De Vliet (Eds.), *Using conflict in organizations* (pp. 10–22). Thousand Oaks, CA: Sage. doi:10.4135/9781446217016.n2
- De Dreu, C. K. W. (2011). Conflict at work: Basic principles and applied issues. In S. Zedeck (Ed.), *APA Handbook of industrial and organizational psychology* (Vol. 3, pp. 461–493). Washington, DC: APA. doi:10.1037/12171-013
- DeShon, R. P., Kozlowski, S. W. J., Schmidt, A. M., Milner, K. R., & Wiechmann, D. (2004). A multiple-goal, multilevel model of feedback effects on the regulation of individual and team performance. *Journal of Applied Psychology, 89*, 1035–1056. doi:10.1037/0021-9010.89.6.1035
- Deutsch, M. (1973). *The resolution of conflict: Constructive and destructive processes*. New Haven, CT: Yale University Press. doi:10.1177/000276427301700206
- Deutsch, M. (2006). Cooperation and competition. In M. Deutsch, P. T. Coleman, & E. C. Marcus (Eds.), *The handbook of conflict resolution: Theory and practice* (2nd ed., pp. 23–42). San Francisco, CA: Wiley
- de Wit, F. R. C., Greer, L. L., & Jehn, K. A. (2012). The paradox of intragroup conflict: A meta-analysis. *Journal of Applied Psychology, 97*, 360–390. doi:10.1037/a0024844
- Drach-Zahavy, A., & Somech, A. (2001a). Understanding team innovation: The role of team processes and structures. *Group Dynamics: Theory, Research and Practice, 5*, 111–123. doi:10.1037/1089-2699.5.2.111
- Drach-Zahavy, A., & Somech, A. (2001b). Can we win them all? Benefits and costs of structure and flexible innovation implementations. *Group Dynamics: Theory, Research and Practice, 25*, 217–234. doi:10.1002/job.239
- Driskell, J. E., Goodwin, G. F., Salas, E., & O'Shea, P. G. (2006). What makes a good team player? Personality and team effectiveness. *Group Dynamics: Theory, Research, and Practice, 10*, 249–271. doi:10.1037/1089-2699.10.4.249
- Edmondson, A. C. (1999). Psychological safety and learning behavior in work teams. *Administrative*

- Science Quarterly*, 44, 350–383. doi:10.2307/2666999
- Emmons, R. A. (1987). Narcissism: Theory and measurement. *Journal of Personality and Social Psychology*, 52, 11–17. doi:10.1037/0022-3514.52.1.11
- Goffin, R. D., Gellatly, I. R., Paunonen, S. V., Jackson, D. N., & Meyer, J. P. (1996). Criterion validation of two approaches to performance appraisal: The behavioral observation scale and the relative percentile method. *Journal of Business and Psychology*, 11, 23–33. doi:10.1007/BF02278252
- Greer, L. L., Jehn, K. A., & Mannix, E. A. (2008). Conflict transformation: A longitudinal investigation of the relationships between different types of intragroup conflict and the moderating role of conflict resolution. *Small Group Research*, 39, 278–302. doi:10.1177/1046496408317793
- Hackman, J. R. (1987). The design of work teams. In J. W. Lorsch (Ed.), *Handbook of organizational behavior* (pp. 315–342). Englewood Cliffs, NJ: Prentice Hall.
- Hackman, J. R. (2002). *Leading teams: Setting the stage for great performances*. Boston, MA: Harvard Business School Press.
- Hayes, A. R. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guilford Press.
- Hodson, G., Hogg, S. M., & MacInnis, C. C. (2009). The role of “dark personalities” (Narcissism, Machiavellianism, Psychopathy), big five personality factors and ideology in explaining prejudice. *Journal of Research in Personality*, 43, 686–690. doi:10.1016/j.jrp.2009.02.005
- Humphrey, S. E., Hollenbeck, J. R., Meyer, C. J., & Ilgen, D. R. (2007). Trait configurations in self-managed teams: A conceptual examination of the use of seeding for maximizing and minimizing trait variance in teams. *Journal of Applied Psychology*, 92, 885–892. doi:10.1037/0021-9010.92.3.885
- Jehn, K. A. (1995). A multimethod examination of the benefits and detriments of intragroup conflict. *Administrative Science Quarterly*, 40, 256–282. doi:10.2307/2393638
- Jehn, K. A., Greer, L., Levine, S., & Szulanski, G. (2008). The effects of conflict types, dimensions, and emergent states on group outcomes. *Group Decision and Negotiation*, 17, 465–495. doi:10.1007/s10726-008-9107-0
- Johnson, D. W. (2003). Social interdependence: Interrelationships among theory, research and practice. *American Psychologist*, 58, 934–945. doi:10.1037/0003-066X.58.11.934
- Johnson, D. W., & Johnson, R. T. (1989). *Cooperation and competition: Theory and research*. Edina, MN: Interaction Book.
- Johnson, J. W. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research*, 35, 1–19. doi:10.1207/S15327906MBR3501\_1
- Johnson, J., & LeBreton, J. M. (2004). History and use of relative importance indices in organizational research. *Organizational Research Methods*, 7, 238–257. doi:10.1177/1094428104266510
- Jonason, P. K., & Krause, L. (2013). The emotional deficits associated with the Dark Triad traits: Cognitive empathy, affective empathy, and alexithymia. *Personality and Individual Differences*, 55, 532–537. doi:10.1016/j.paid.2013.04.027
- Jonason, P. K., Slomski, S., & Partyka, J. (2012). The dark triad at work: How toxic employees get their way. *Personality and Individual Differences*, 52, 449–453. doi:10.1016/j.paid.2011.11.008
- Jonason, P. K., & Tost, J. (2010). I just cannot control myself: The dark triad and self-control. *Personality and Individual Differences*, 49, 611–615. doi:10.1016/j.paid.2010.05.031
- Jonason, P. K., & Webster, G. D. (2010). The dirty dozen: A concise measure of the dark triad. *Psychological Assessment*, 22, 420–432. doi:10.1037/a0019265
- Jonason, P. K., & Webster, G. D. (2012). A protean approach to social influence: Dark Triad personalities and social influence tactics. *Personality and Individual Differences*, 52, 521–526. doi:10.1016/j.paid.2011.11.023
- Jonason, P. K., Webster, G. D., Schmitt, D. P., Li, N. P., & Crysel, L. (2012). The antihero in popular culture: Life history theory and the dark personality traits. *Review of General Psychology*, 16, 192–199. doi:10.1037/a0027914
- Jones, D. N., & Paulhus, D. L. (2011). The role of impulsivity in the dark triad of personality. *Personality and Individual Differences*, 51, 679–682. doi:10.1016/j.paid.2011.04.011
- Jordan, P. J., & Troth, A. C. (2004). Managing emotions during team problem solving: Emotional intelligence and conflict resolution. *Human Performance*, 17, 195–218. doi:10.1207/s15327043hup1702\_4
- Judd, C. M., & Kenny, D. A. (2010). Data analysis in social psychology: Recent and recurring issues. In S. T. Fiske, D. T. Gilbert, & L. Gardner (Eds.), *Handbook of social psychology* (5th ed., Vol. 1, pp. 115–139). Mahwah, NJ: Wiley.
- Judge, T. A., & LePine, J. A. (2007). The bright and dark sides of personality: Implications for personnel selection in individual and team contexts. In J. Langen-Fox, C. L. Cooper, & R. J. Klimoski (Eds.), *Research companion to the dysfunctional workplace: Management challenges and symptoms*



- (pp. 332–355). London, UK: Edward Elgar Publishing Ltd. doi:10.4337/9781847207081.00028
- Katz, D., & Kahn, R. L. (1978). *The social psychology of organizations* (2nd ed.). New York, NY: Wiley and Sons.
- Korsgaard, M. A., Jeong, S. S., Mahony, D. M., & Pitariu, A. H. (2008). A multilevel view of intragroup conflict. *Journal of Management*, *34*, 1222–1252. doi:10.1177/0149206308325124
- Kozlowski, S. W. J., & Bell, B. S. (2003). Work groups and teams in organizations. In W. C. Borman, D. R. Ilgen, & R. J. Klimoski (Eds.), *Handbook of psychology* (Vol. 12): *Industrial and organizational psychology* (pp. 333–375). New York, NY: Wiley. doi:10.1002/9781118133880.hop212017
- 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: Foundations, extensions, and new directions* (pp. 3–90). San Francisco, CA: Jossey-Bass.
- Kuhn, T., & Poole, M. S. (2000). Do conflict management styles affect group decision making? Evidence from a longitudinal field study. *Human Communication Research*, *26*, 558–590. doi:10.1111/j.1468-2958.2000.tb00769.x
- Lee, K., & Ashton, M. C. (2005). Psychopathy, Machiavellianism, and Narcissism in the five-factor model and the HEXACO model of personality structure. *Personality and Individual Differences*, *38*, 1571–1582. doi:10.1016/j.paid.2004.09.016
- LePine, J. A. (2003). Team adaptation and postchange performance: Effects of team composition in terms of members' cognitive ability and personality. *Journal of Applied Psychology*, *88*, 27–39. doi:10.1037/0021-9010.88.1.27
- LePine, J. A., Buckman, B. R., Crawford, E. R., & Methot, J. R. (2011). A review of research on personality in teams: Accounting for pathways spanning levels of theory and analysis. *Human Resource Management Review*, *21*, 311–330. doi:10.1016/j.hrmr.2010.10.004
- LePine, J. A., Hollenbeck, J. R., Ilgen, D. R., & Hedlund, J. (1997). Effects of individual differences on the performance of hierarchical decision-making teams: Much more than g. *Journal of Applied Psychology*, *82*, 803–811. doi:10.1037/0021-9010.82.5.803
- Levenson, M. R., Kiehl, K., & Fitzpatrick, C. M. (1995). Assessing psychopathic attributes in a non-institutionalised population. *Journal of Personality and Social Psychology*, *68*, 151–158. doi:10.1037/0022-3514.68.1.151
- Lewicki, R. J., Weiss, S. E., & Lewin, D. (1992). Models of conflict, negotiation, and third party intervention: A review and synthesis. *Journal of Organizational Behavior*, *13*, 209–252. doi:10.1002/job.4030130303
- McGrath, J. E. (1984). *Groups: Interaction and performance*. Englewood Cliffs, NJ: Prentice Hall.
- McHoskey, J. W., Worzel, W., & Szyarto, C. (1998). Machiavellianism and psychopathy. *Journal of Personality and Social Psychology*, *74*, 192–210. doi:10.1037/0022-3514.74.1.192
- O'Neill, T. A., & Allen, N. A. (2011). Personality and the prediction of team performance. *European Journal of Personality*, *25*, 31–42. doi:10.1002/per.769
- O'Neill, T. A., & Allen, N. A. (2012). Team meeting attitudes: Conceptualization and investigation of a new construct. *Small Group Research*, *43*, 186–210. doi:10.1177/1046496411426485
- O'Neill, T. A., Allen, N. A., & Hastings, S. E. (2013). Examining the “pros” and “cons” of team conflict: A team-level meta-analysis of task, relationship, and process conflict. *Human Performance*, *26*, 236–260. doi:10.1080/08959285.2013.795573
- Paulhus, D. L., & Williams, K. M. (2002). The dark triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*, *36*, 556–563. doi:10.1016/S0092-6566(02)00505-6
- Paunonen, S. V. (2002). Design and construction of the Supernumerary Personality Inventory. *Research Bulletin* (Vol. 763). London, Ontario: University of Western Ontario.
- Paunonen, S. V., Haddock, G., Forsterling, F., & Keinonen, M. (2003). Broad versus narrow personality measures and the prediction of behavior across cultures. *European Journal of Personality*, *17*, 413–433. doi:10.1002/per.496
- Paunonen, S. V., & Jackson, D. N. (2000). What is beyond the Big Five? Plenty! *Journal of Research in Personality*, *68*, 821–835. doi:10.1111/1467-6494.00117
- Peterson, R. S., & Behfar, K. J. (2003). The dynamic relationship between performance feedback, trust, and conflict in groups: A longitudinal study. *Organizational Behaviour and Human Decision Processes*, *92*, 102–112. doi:10.1016/S0749-5978(03)00090-6
- Petrides, K. V., Vernon, P. A., Schermer, J. A., & Veselka, L. (2011). Trait emotional intelligence and the dark triad traits of personality. *Twin Research and Human Genetics*, *14*, 35–41. doi:10.1375/twin.14.1.35
- Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments & Computers*, *36*, 717–731. doi:10.3758/BF03206553

- Rahim, M. A. (1983). A measure of styles of handling interpersonal conflict. *Academy of Management Journal*, *26*, 368–376. doi:10.2307/255985
- Raskin, R., & Terry, H. (1988). A principle components analysis of the narcissistic personality inventory and further evidence of its construct validity. *Journal of Personality and Social Psychology*, *54*, 890–902. doi:10.1037/0022-3514.54.5.890
- Rauthmann, J. F. (2012). The Dark Triad and interpersonal perception: Similarities and differences in the social consequences of Narcissism, Machiavellianism, and Psychopathy. *Social Psychological and Personality Science*, *3*, 487–496. doi:10.1177/1948550611427608
- Robey, D., & Farrow, D. L. (1982). User involvement in information system development: A conflict model and empirical test. *Management Science*, *28*, 73–85. doi:10.1287/mnsc.28.1.73
- Robey, D., Farrow, D. L., & Franz, C. R. (1989). Group process and conflict in system development. *Management Science*, *35*, 1172–1191. doi:10.1287/mnsc.35.10.1172
- Ross, S. R., Bye, K., Wrobel, T. A., & Horton, R. S. (2008). Primary and secondary psychopathic characteristics and the schedule for non-adaptive and adaptive personality (SNAP). *Personality and Individual Differences*, *45*, 249–254. doi:10.1016/j.paid.2008.04.007
- Rubin, J. Z., Pruitt, D. G., & Kim, S. H. (1994). *Social conflict: Escalation, stalemate, and settlement* (2nd ed.). New York, NY: McGraw-Hill.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420–428. doi:10.1037/0033-2909.86.2.420
- Society for Industrial-Organizational Psychology. (2003). *Principles for the validation and use of personnel selection procedures* (4th ed.). Bowling Green, OH: SIOP.
- Sundstrom, E., McIntyre, M., Halfhill, T., & Richards, H. (2000). Work groups: From the Hawthorne studies to work teams of the 1990s and beyond. *Group Dynamics: Theory, Research, and Practice*, *4*, 44–67. doi:10.1037//1089-2699.4.1.44
- Tjosvold, D. (1998). Cooperative and competitive goal approach to conflict: Accomplishments and challenges. *Applied Psychology*, *47*, 285–342. doi:10.1111/j.1464-0597.1998.tb00025.x
- Wageman, R., & Gordon, F. M. (2005). As the twig is bent: How group values shape emergent task interdependence in groups. *Organization Science*, *16*, 687–700. doi:10.1287/orsc.1050.0146
- Wai, M., & Tiliopoulos, N. (2012). The affective and cognitive empathic nature of the dark triad of personality. *Personality and Individual Differences*, *52*, 794–799. doi:10.1016/j.paid.2012.01.008
- Wall, J. A., & Callister, R. R. (1995). Conflict and its management. *Journal of Management*, *21*, 515–558. doi:10.1177/014920639502100306
- Wegge, J., van Dick, R., Fisher, G. K., West, M. A., & Dawson, J. F. (2006). A test of basic assumptions of affective events theory (AET) in call centre work. *British Journal of Management*, *17*, 237–254. doi:10.1111/j.1467-8551.2006.00489.x
- Weiss, H. M., & Cropanzano, R. (1996). Affective events theory: A theoretical discussion of the structure, causes and consequences of affective experiences at work. *Research in Organizational Behavior*, *18*, 1–74.

Received May 7, 2013

Revision received November 14, 2013

Accepted November 20, 2013 ■

# The Impact of Group Norms and Behavioral Congruence on the Internalization of an Illegal Downloading Behavior

Sophie Sansfaçon and Catherine E. Amiot  
Université du Québec à Montréal

This experimental study ( $n = 115$ ) investigates how group norms and individuals' congruence with these norms predicted internalization (i.e., self-determination) of an illegal behavior. We manipulated ingroup norms in favor of versus against illegal downloading of software, and assessed participants' behavioral intentions, their motivations for emitting this specific behavior, and how they subjectively experienced the experimental situation (compartmentalizing their illegal behavior; feeling conflicted). Participants were more likely to endorse the behavior that was normative in their ingroup. Also, participants were more self-determined to engage in the behavior when they intended to behave congruently with a norm against illegal downloading, and when they intended to behave incongruently with a norm in favor of illegal downloading. Participants who did not intend to engage in illegal downloading felt less compartmentalized and conflicted than participants who intended to engage in illegal downloading. Results are discussed in light of self-determination theory (Deci & Ryan, 2000) and social identity theory (Tajfel & Turner, 1986).

**Keywords:** dissent, group norms, illegal behavior, norm and behavior congruence, self-determination

What drives people to behave illegally, such as by stealing or cheating? One could say that the reasons why people act illegally are instrumental, and serve to access valued resources (Lee & Allen, 2002) or gain social approval (Baumeister & Vohs, 2004). For example, students may have an *external motivation* to illegally download a computer software from the Internet because they need that program to succeed in their studies. Other students may engage in the illegal behavior in a more *internalized manner*, because they feel that it is right for students not to pay for computer software because they don't have a salary yet, and that the

software companies are already making a lot of money. Still, we can wonder about the variables that facilitate the internalization of such illegal behavior within the self. Previous studies have demonstrated that group norms influence not only the emission of negative behaviors (Frone & Brown, 2010; Kura, Shamsudin, & Chauhan, 2013; Robinson & O'leary-Kelly, 1998), but also the internalization of those behaviors (Amiot, Sansfaçon, & Louis, 2013; Amiot, Sansfaçon, Louis & Yelle, 2012). When perceiving one's group norm as encouraging a behavior, it is possible that those behaviors become internalized by individual group members, such that they are part of their self (Haslam & Reicher, 2007) and are engaged in out of pleasure or because they align with one's values (Amiot, Sansfaçon, & Louis, 2013; *in press*). In the current research, we aim to test, in an experimental context, whether a situationally salient norm influences the immediate *internalization* of the normative behaviors.

In fact, past studies have showed that when an ingroup is important to the individual, the norms of this group can become internalized and impact significantly on the behaviors of individual group members (e.g., Fielding, Mc-

---

Sophie Sansfaçon and Catherine E. Amiot, Département de Psychologie, Université du Québec à Montréal, Montréal, Québec, Canada.

This research was facilitated thanks to grants from the Social Sciences and Humanities Research Council of Canada (SSHRC) and from the Fond pour la recherche en santé-Québec (FRSQ) to Catherine E. Amiot.

Correspondence concerning this article should be addressed to Sophie Sansfaçon, Département de psychologie, Université du Québec à Montréal, C.P. 8888, Succ. Centre-Ville, Montréal, PQ, H3C 3P8. E-mail: [cpf.sansfaçon@gmail.com](mailto:cpf.sansfaçon@gmail.com)

Donald, & Louis, 2008; Terry & Hogg, 1996). Relying on self-determination theory (Deci & Ryan, 2000) and intergroup theories based on the social identity approach (Tajfel & Turner, 1986), the first goal of the current study is to investigate the internalization of a particular illegal behavior, namely, the illegal downloading of computer software. This behavior is private, anonymous, and relatively common (Belgian Anti-Piracy Federation, 2009); it was considered relevant to study in relation to social identification and internalization processes given that social influence and norms operate strongly under conditions of isolation and anonymity (Spears, Postmes, Lea, & Wolbert, 2002). We expected this illegal behavior to be internalized when social norms and cues are salient.

Our second goal is to investigate how participants subjectively feel when they behave in line with group norms or not. To this aim, we assessed how engaging in an illegal behavior is compartmentalized and to what extent it can cause conflict within the self. We also explore how participants subjectively perceive how different sources of influence guided their behavior in the experimental setting. Doing so opens the door to a new subjective perspective on group norms that departs from traditional norms research, which has typically looked at objective (e.g., manipulated) sources of normative influence that operate outside of individuals' awareness (Vorauer & Miller, 1997).

### Social Norms Promote Internalization

Social norms are useful guides for what is considered appropriate behavior in group settings. Norms represent what is done and prescribed by the other group members (Jacobson, Mortensen, & Cialdini, 2011). Despite being potentially coercive (Turner, 1991), norms can also be internalized by group members and come to be followed freely and out of choice (Amiot, Sansfaçon, & Louis, 2013). Internalization refers to the process by which an individual is taking in the values and behaviors of a group and integrating them within their sense of self, by making them their own (Kelman, 1961). This process operates beyond acting for personal gain, or to avoid the negative consequences that derive from violating the norm (Kerr, Garst, Kiehle, & Harris, 1997). From a

social identity perspective, when an individual identifies with a group, the group identity becomes a part of the self, and so do the norms of this group (Turner, 1991). Whether norms are followed or not also depends on the social context: when a group membership is salient in a social context and this membership is important to the person's self-definition, group norms will become activated and will guide one's behavior (Smith & Louis, 2008; Terry & Hogg, 1996). This internalization is also facilitated by intra-group processes, including communication that takes place within the ingroup (Kerr & Kaufman-Gilliland, 1994) and consensualization processes (Turner, 1991).

For example, at a hockey game, if a hockey fan identifies strongly as a fan of the team and perceives that his or her team group is encouraging him/her in engaging in a harmful behavior against fans of other team, s/he will tend to engage in the normative behavior and internalize this behavior to a greater extent (Amiot, Sansfaçon, & Louis, *in press*). Importantly, the intergroup perspectives propose that the internalization process should take place regardless of whether the normative group behaviors are prosocial or potentially harmful (Turner, 1991) – such as whether behaviors are illegal or not.

### Self-Determination Theory

Previous studies from the organizational literature have confirmed that motivations to engage in illegal behaviors in the workplace can be diverse, from thrill seeking to role modeling, and to feelings that such behaviors allow to restore justice (Robinson & Greenberg, 1998). This suggests that negative behavior can be emitted out of different motives. Herein we focus on motives that range from being driven by external sources, to behaviors that are followed for more internal reasons (Amiot, Sansfaçon, & Louis, 2013). To concretely capture the possible motivations for people's behaviors and the extent to which these behaviors—including normative behaviors—are internalized, the present research relies on self-determination theory (SDT; Deci & Ryan, 2000). SDT is a motivational theory that has been applied mostly to understanding the motivations that underlie constructive social behaviors (e.g., volunteering; Millette & Gagné, 2008). SDT proposes that when psychological

needs (autonomy, relatedness, competence) are satisfied by the social environment, behaviors tend to be emitted out of self-determined motives (internalized), and individuals report higher well-being (Lynch, La Guardia, & Ryan, 2009). For SDT theorists, motivations vary along a continuum ranging from reasons that are not internalized and are not considered as part of the self (*non-self-determined*), to reasons that are considered internalized and part of the self (*self-determined*). More precisely, the non-self-determined motivations include amotivation, external regulation, and introjected regulation. Amotivation represents a total absence of motivation. External regulation is present when an individual acts out of external contingencies, such as to obtain rewards or to flee punishment. Introjected regulation occurs when the individual is acting out of internal pressure, so as not to feel bad or guilty.

The more internalized and self-determined motivations include identified regulation, integrated regulation, and intrinsic motivation. Identified regulation is present when individuals feel that engaging in the behavior is important for them and allows them to reach valued goals. Integrated regulation is present when individuals act because the behavior is coherent with their personal objectives and values. Finally, intrinsic motivation occurs when individuals behave out of pleasure and for the satisfaction of engaging in the activity itself.

These motivations have been explored in various contexts (Vallerand, 1997), and are beginning to be applied to group settings (e.g., Amiot & Sansfaçon, 2011). In the current research, we employ the SDT framework not only to capture internalization, but also because this framework proposes that socially harmful behaviors will be more difficult to internalize in the self (Deci & Ryan, 2000). Whereas the intergroup approaches do not make distinctions between socially harmful and beneficial group norms with respect to their capacity to become internalized in the self (Turner, 1991), SDT clearly puts limits to this capacity for internalization by suggesting that potentially harmful social behaviors are fuelled by a *lack* of autonomy and self-determination, and that such behaviors occur when one's basic psychological needs have been thwarted (Deci & Ryan, 2000). In fact, SDT is a humanistic theory that assumes that humans strive to act effectively and in a proso-

cial manner (Deci & Ryan, 2000). Empirically, studies have found that abusive supervision in the workplace context is linked to organizational deviance, and that this link is mediated by need satisfaction (Lian, Ferris, & Brown, 2012). In the current research, we expected, based on SDT, that harmful behaviors—such as illegal behaviors—should hence trigger compartmentalization and intraindividual conflict and unease given that such behaviors clash with the person's true goals and aims.

Prior studies confirmed that prosocial (rather than harmful) norms are easier to internalize and that group norms and individuals' behavioral congruence with the norms do have an impact on the internalization of a specific normative behavior (Amiot et al., 2012). These studies tested competing hypotheses based on SIT versus SDT about the possibility of group norms to be internalized and endorsed out of self-determination. The findings tended to support SDT's perspective: When an important ingroup was encouraging a pro-social behavior, such as distributing resources equally between groups (prosocial parity norm), and individuals were behaving congruently with this norm, this behavior was more strongly internalized and self-determined (Amiot et al., 2012). The reverse pattern was observed for harmful norms in favor of discrimination: When one's ingroup encouraged discrimination, it is individuals who behaved *incongruently* with this norm (by engaging instead in parity) who were more self-determined. These results suggest that ingroup norms do have an impact on the internalization of the behavior, and that engaging in a potentially harmful social behavior tends to be more difficult to internalize in the self, even if this behavior is proscribed by ingroup norms. As well, and in line with Packer (2008), participants who went against a socially harmful norm in favor of discrimination by engaging in a pro-social behavior instead may have drawn their motivation from particularly internalized values. Risking to go against the ingroup norms can be costly for the individual with regard of his or her membership within a group, which suggests that deviant behaviors are engaged in out of deeply held values and beliefs.

The current study will test whether the same results are observable when we manipulate norms that concern an illegal behavior, namely the illegal downloading and use of

computer software. This type of behavior is considered to be socially harmful from a legal and a moral point of view. The law considers pirated software downloading as theft (Copyright Board of Canada, 2009). Downloading software illegally does not recognize the work conducted by software designers and computer companies. Moreover, not paying for a product does not encourage software development and increases the costs of software for buyers (BAF, 2009). We expected that this type of behavior will be possible to change via the manipulation of situational norms given that it is an anonymous behavior (Spears et al., 2002) that is considered quite benign in terms of its consequences (Aguiar & Martens, 2013). Furthermore, different studies conducted among university studies report diverging statistics regarding the frequency of illegal downloading behaviors: For example, whereas a Belgian survey established that 90% of students engage in the illegal downloading of software (BAF, 2009), a French survey revealed that about 40% of students engage in such behavior (JDN, 2004). This divergence suggests that norms regarding the prevalence of illegal downloading are somewhat vague, and that the norms manipulated in this context may hence be relied on more strongly as a means to resolve this ambiguity and vagueness (Smith, Hogg, Martin, & Terry, 2007). This allows us to expect that the norm manipulation used in the current study will have a potent impact on participants' downloading behaviors and on their internalization of these behaviors.

### Subjective Processes

The experiment will also explore three new variables that capture individual group members' subjective experiences of the norm internalization process. More specifically, we investigate how individual group members compartmentalize their illegal behaviors within themselves, the level of internal pressure or conflict they subjectively feel when they intend to engage or not in a normative illegal behavior, and which specific sources of normative influence they believe are influencing them when intending to engage (or not) in illegal behaviors.

### Compartmentalization

Compartmentalization is a term used when referring to distinct parts of the self that are kept separated from each other (Ryan & Deci, 2003). Compartmentalization of behaviors occurs when individuals behave in a way that fits with the social identity and the social norms that are salient in one specific social context, but without linking back this behavior to other life contexts or to their own personal values. For example, a prisoner can act in an illegal manner when s/he is in the prison context and his or her prisoner identity is salient, but can follow laws and regulations when s/he is in another group, her/his family for instance (Tripp, 2009). In this case, the illegal behavior does not represent the individual as a whole, but represents the person only in one particular context. Compartmentalization should be experienced to a greater extent when individuals intend to engage in illegal behaviors, because these kinds of behaviors are generally disapproved of by society, and need to be kept in a distinct and hermetic part of the self so as not to contaminate their entire sense of self (Ryan & Deci, 2003). We expect that individuals who intend to act illegally (who intend to behave congruently with a downloading norm or incongruently with a no-downloading norm) will report higher compartmentalization than those who intend to act legally.

### Intraindividual Conflict

Similarly, conflict arises when an individual feels that there is a discrepancy between different parts of him or herself, such as between diverging norms or values (Benet-Martínez & Haritatos, 2005). Conflict may arise when the individual acts in a way that does not represent his or her own personal values and attitudes, or when following group norms that are *proscribed* by another of his or her social groups. Because acting illegally is considered to be a socially harmful behavior and also based on SDT postulates, we hypothesize that intraindividual conflict will be positively related to compartmentalization. Moreover, we expect that individuals who intend to act illegally (who intend to behave congruently with a downloading norm or incongruently with a no-downloading norm) will feel more conflict than those who act legally.

## Perceived Influence

According to some researchers, the social influence process is an implicit, unconscious mechanism (Vorauer & Miller, 1997), which also implies that the subjective perception of norms can be biased (Baer, Stacy & Larimer, 1991). Still, we can wonder how individuals perceive what has influenced them in a specific situation, such as group norms or personal values (Van Bavel, Packer, & Cunningham, 2008). This study will hence explore how participants subjectively perceive different possible sources of influence—namely the situation, their own personal values and sense of morality, and in-group norms—and how they consider that these sources guided their behavior in the experimental context. Exploring these subjective perceptions is important given that multiple sources of influence have the potential to guide people's behaviors in a specific context (i.e., attitudes and social norm; Ajzen, 2001) and that we wish to capture what influences people's illegal actions, from their own point of view.

The perception of what influenced one's own behavior is probably a process that relates to the compartmentalization process: A person who compartmentalizes his or her behavior associates and restricts this behavior to a particular life context. In this sense, the more a person compartmentalizes his or her behavior, the more likely we expect that he or she will perceive that it is a specific group or the situation itself (rather than his or her own morality or personal values) that influenced the behavior. We hence expect that the different sources of influence will relate differently with compartmentalization.

## The Present Research

Following past studies (Amiot et al., 2012), the first objective of this research is to test whether a “prodownloading norm” (i.e., a norm in favor of illegal downloading) versus a “no-downloading norm” (i.e., a norm that is not against illegal downloading) influences participants' behavioral intentions to engage in illegal downloading behaviors. The second objective is to investigate how individuals' congruence/incongruence with these norms influences their self-determination (as a proxy for internalization) for engaging in this behavior. A third objective is to capture participants' subjective

experiences in the norm internalization process, and more specifically their feelings of compartmentalization and intraindividual conflict, along with the sources of influence that they perceived as guiding their behavior in the experiment. By doing so, the present study goes a step further in a line of research bridging SDT and intergroup theories. By testing these hypotheses with a potentially malleable illegal behavior, and investigating these additional subjective processes, this study will allow a broader look at the consequences of engaging in and following socially “harmful” norms. Moreover, whereas other studies have explored the internalization of negative social norms within correlational design in real life contexts (e.g., discrimination among hockey fans: Amiot, Sansfaçon, & Louis, in press), the present study will test whether a situational group norm will influence the immediate *internalization* of such behaviors, in an experimental context.

To reach these goals, we present an experimental study in which we manipulated group norms and assessed participants' congruence with these norms. In line with past studies on norm internalization (Amiot et al., 2012), we decided to manipulate the *descriptive* norm here (individual's perception of what other group members do), instead of *injunctive* norms (perception of what group members encouraged individual to do; Reno, Cialdini, & Källgren, 1993; Smith & Louis, 2008; Smith et al., 2008).

Based on prior research, the following hypotheses were made: (H1) Ingroup norms will influence participants' behavioral intentions: Participants exposed to a prodownloading ingroup norm will themselves report stronger intentions to engage in illegal prodownloading in comparison with participants exposed to a no-downloading ingroup norm. In line with prior studies (Amiot et al., 2012), H2a predicts that participants who act congruently with the no-downloading norm (and hence who do not intend to behave illegally themselves) will be more self-determined than those who behave incongruently with this norm. Among participants in the prodownloading condition, those who intend to behave *incongruently* with the norm should report higher self-determination than those who intend to behave in line with this norm (H2b). H3 and H4 predict that participants will report higher compartmentalization and more conflict when they intend to act illegally,

that is, both when they plan to act congruently with a prodownloading norm or incongruently with a no-downloading norm. Based on a subjective approach to group norms, H5 predicts that compartmentalization will be positively related to intraindividual conflict. We also anticipate that the more a person compartmentalizes his or her behavior, the more likely we expect that he or she will perceive that it is a specific group or the situation itself (rather than his or her own morality or personal values) that influenced his or her behavior (H6). Finally, given the lack of prior studies on how individuals subjectively perceive different sources of influence, we will explore how these sources of influence are perceived across experimental conditions.

## Method

### Participants and Design

Participants were 123 students enrolled at a French-speaking university in Montréal. The study employed a between-participants design where we manipulated social identity salience (personal vs. student identity; Haslam, Oakes, Reynolds, & Turner, 1999)<sup>1</sup> and ingroup norms (prodownloading vs. no-downloading of computer software). No participant had suspicions about what the goal of the study was, but 8 participants did not understand the ingroup norm manipulation correctly and were excluded from the sample. The final sample ( $N = 115$ ) included 90 females and 24 males (1 participant did not report their gender), who were on average 22.50 years old ( $SD = 4.74$ ).

### Procedure

Participants were presented with information about the general context of illegal software downloading among university students. They were told that studying at University requires buying and downloading costly computer programs, and that some of the students may illegally download those programs. In line with past research (Smith & Louis, 2008), participants were then informed that a survey had been conducted among students of their University (UQAM) to gather their opinions and behaviors regarding illegal downloading. Bar graphs were presented and illustrated the results observed

when asking students the following question: *Do you personally illegally download and use computer software?* For participants in the downloading norm condition, the graphs showed that 76% of students said that they were engaging in illegal downloading while 24% were not. For participants in the no-downloading norm condition, these percentages were reversed.

To strengthen the norm manipulation further, participants read a series of five opinion statements about the target issue of computer software downloading, ostensibly from five students from their program who had participated in the survey before them. The goal with this manipulation was to provide participants with further ideological justifications for endorsing the norm (Amiot & Bourhis, 2003). These statements indicated that four of five students from their program either were engaging in the behavior themselves (or not, depending on the experimental condition), and that one of five was neutral. As a manipulation check, participants were asked whether the graphs and comments showed that the majority of the students at their university were engaging in the behavior that involves illegally downloading computer software or not.

<sup>1</sup> *Identity Salience Manipulation.* In line with past research (Haslam et al., 1999) and to test whether identity salience accentuates the internalization of social norms (Smith & Louis, 2008), a social identity manipulation was first presented. This identity salience manipulation was based on Haslam et al. (1999): half of the participants were primed with their personal identity, and the other half with their social identity as a student of their university. Participants in the social identity condition were asked to answer four questions pertaining to their student identity. The same questions were asked to participants in the personal identity condition, but concerning themselves as individuals. Following the manipulation, two manipulation check questions were asked: *To what extent it is easy for you to describe yourself as a UQAM student?* (high identification condition:  $M = 4.25$ ;  $ST = 1.50$ ; low identification condition:  $M = 4.55$ ;  $SD = 1.35$ ); *To what extent your UQAM student identity is important for you?* (high identification condition:  $M = 5.09$ ;  $ST = 1.35$ ; low identification:  $M = 5.20$ ;  $SD = 1.26$ ), and answered on a 1 to 7 Likert scale, ranging from *Not at all* to *Extremely*. To assess the impact of the identity salience manipulation, ANOVAs were conducted on the two manipulation check questions. No significant difference emerged between the two conditions on the two questions. The identity salience manipulation did not show any main or interactive effects with the norms manipulation on self-determination index,  $t(112) = 1.127$ ,  $p = .262$ ;  $t(112) = .497$ ,  $p = .633$ .



## Measures

**Behavioral intentions.** To assess participants' behavioral intentions, participants were asked whether they would illegally download and use software if their program of study requires it (*yes/no*). Answers to this question were coded to indicate that participants' behavior was either congruent or incongruent with their ingroup norm.

**Motivation for emitting the behavior displayed.** Participants were then asked to think of the reasons why they intend to engage in this behavior. They were then provided with 24 items based on previously validated SDT scales (Guay, Mageau, & Vallerand, 2003). These items measured amotivation (*I am not really sure that this behavior is really worth it*;  $\alpha = .74$ ), external regulation (*because it allows me to do what is expected from me*;  $\alpha = .71$ ), introjected regulation (*because I would have felt ashamed if I had behave in another way*;  $\alpha = .90$ ), identified regulation (*because I believe that it is important to behave in that way*;  $\alpha = .65$ ), integrated regulation (*because this manner of behaving is in line with my values*;  $\alpha = .92$ ), and intrinsic motivation (*because I derived satisfaction from behaving this way*;  $\alpha = .93$ ). Each of the six motivation types included four items. An index of self-determined motivation was created by using the following formula:  $(3 \times \text{intrinsic motivation} + 2 \times \text{integrated regulation} + 1 \times \text{identified regulation}) - (1 \times \text{introjected regulation} + 2 \times \text{external regulation} + 3 \times \text{amotivation})$ ; see Vallerand & Bissonnette, 1992). It is common for SDT researchers to compute an overall self-determination index (Ryan & Connell, 1989; Vallerand, 1997; Vallerand & Bissonnette, 1992). Greater scores on this index represent higher self-determination motives relative to non-self-determination motives. For this measure and the following ones, 1 (*not at all*) to 7 (*extremely*) Likert scales were used (the index ranges from  $-36$  up to  $36$ ).

**Compartmentalization.** Based on Ryan and Deci's (2003) conceptualization of compartmentalization, a 5-item scale was created to measure the extent to which participants' illegal downloading behavior in the context of this study was compartmentalized in their sense of self (*My downloading behaviors do*

*not represent who I am as a person; My downloading behaviors are only emitted in the current context; My downloading behaviors are specific to a particular context of my life; My downloading behaviors represent only a small part of who I am as a person; My downloading behaviors are associated with a particular situation in my life*;  $\alpha = .84$ ).

**Intraindividual conflict.** To assess the conflict between the behavior participants intend to emit and their personal values, five items were adapted from the Benet-Martínez and Haritatos' (2005) cultural conflict scale (e.g., *My behaviors pertaining to computer downloading are in conflict with my personal values; I feel torn between my downloading behaviors and my personal values*;  $\alpha = .86$ ).

**Perceived influence.** To assess the perceived influence of different possible sources, we asked participants to what extent five sources influenced their downloading behavior in the experiment: The situation, their personal values, their morality, the software companies, and other UQAM students.

## Results

### Manipulations Checks: Norm and Behaviors Displayed

To assess the impact of the ingroup norm manipulation on participants' behavioral intentions, chi-square analyses were conducted. A main effect of the ingroup norm emerged,  $\chi^2(df = 1, n = 114) = 10.10, p = .001$ : Participants exposed to the prodowloading norm displayed more downloading intentions in comparison with participants exposed to the no-downloading norm. More specifically, among participants whose ingroup endorses the norm in favor of illegal downloading, 79% ( $n = 44$ ) displayed downloading intentions and 21% ( $n = 12$ ) did not display downloading intentions. Among participants whose ingroup endorsed a no-downloading norm, 50% ( $n = 29$ ) displayed downloading intentions and 50% ( $n = 29$ ) did not display downloading intentions. These results confirm H1: Ingroup norms significantly influenced participants' behavioral intentions.

**Descriptive Statistics and Correlations Among All Variables**

Table 1 displays the correlations between the variables for participants in all conditions. As could be expected because self-determination reflects greater internalization and engaging in behaviors because they reflect one’s choices and values, the self-determination index correlates positively with perceiving that what influenced one in the experiment are their personal values ( $r = .44, p < .01$ ) and morality ( $r = .41, p < .01$ ), but negatively with perception of situational influence ( $r = -.20, p < .05$ ). Moreover, supporting H5, compartmentalization correlates with intraindividual conflict ( $r = .40, p < .01$ ) and with situational influence ( $r = .41, p < .01$ ), which is also coherent with the definition of compartmentalization as taking place when distinct behaviors are kept separated from each other in the self and associated only with specific situations (Ryan & Deci, 2003). Compartmentalization also correlates negatively with the perception that morality influenced one’s behavior ( $r = -.22, p < .05$ ), which is also coherent with the concept and with H6. Finally, and in line with social identity theory, perceiving the students as influential is related to identifying with the student ingroup ( $r = .21, p < .05$ ).

**Motivation for Emitting the Behavior**

To assess the impact of the ingroup norm and the congruence of participants’ intentions relative to this norm on self-determination, a

2 (ingroup norm: prodownloading/no-downloading norm)  $\times$  2 (congruence: participant’s behavioral intentions are congruent/incongruent with the ingroup norm) ANOVA was conducted on the self-determination index. The main effect of the ingroup norm was significant,  $F(1, 112) = 10.67, p < .001, \eta^2 = .09$ : Participants in the no-downloading norm condition were more self-determined ( $M = 12.04; SD = 10.90$ ) than participants in the prodownloading norm condition ( $M = 3.80; SD = 9.16$ ). The main effect of congruence was not significant,  $F(1, 112) = 0.01, p = .923, \eta^2 = .000$ . A significant interaction was observed between ingroup norm and congruence (Figure 1):  $F(1, 112) = 17.295, p < .001, \eta^2 = .14$ ; participants whose behavioral intentions were congruent with the no-downloading norm condition (and who did not engage in illegal downloading) reported greater self-determination ( $M = 8.54; SD = 7.81$ ) than participants whose behavioral intentions were incongruent with this norm (and who engaged in illegal downloading instead;  $M = 1.69; SD = 9.32; F(1, 57) = 8.86, p = .004, \eta^2 = .14$ ). In contrast, participants whose behavioral intentions were incongruent with a prodownloading norm (such that they did not engage in illegal downloading) were more self-determined ( $M = 3.13; SD = 9.80$ ) than participants whose behavioral intentions were congruent with this norm (and who hence engaged in illegal downloading;  $M = -3.40; SD = 4.88, F(1, 55) = 10.02, p = .003, \eta^2 = .16$ ). These findings support H2a and H2b.

Table 1  
*Descriptive Statistics and Correlations Among All Variables*

Variables	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7
1. Index	7.71	8.32							
2. Compartmentalization	3.49	1.76	-.40**						
3. Intraindividual conflict	2.69	1.36	-.50**	.40**					
4. Influence: Situation	5.15	1.57	-.20*	.41**	.17				
5. Influence: Personal values	4.85	1.78	.44**	-.15	-.17	-.08			
6. Influence: Morality	4.56	2.03	.41**	-.22*	-.08	.00	.87***		
7. Influence: Software companies	3.23	2.01	.11	-.05	-.12	.23*	.12	.12	
8. Influence: Students	2.52	1.67	-.04	.07	.11	.16	.14	.42**	
9. Student identification	4.78	1.17	-.06	.15	.16	-.02	-.01	.09	.21*

Note. *M* = mean; *SD* = standard deviation.  
\*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$ .

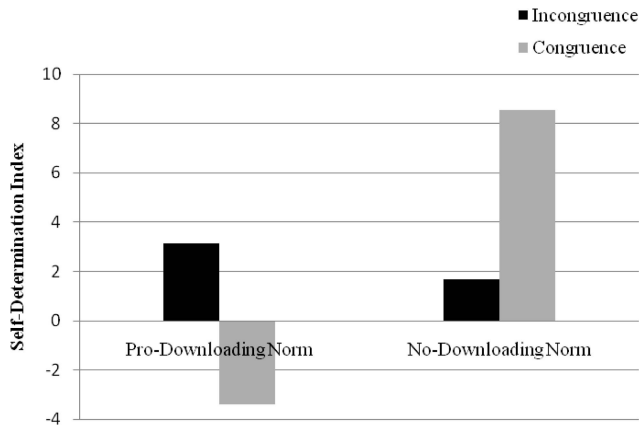


Figure 1. Self-determination as a function of ingroup norm and congruence.

### Compartmentalization

To assess the impact of the ingroup norm and the congruence of participants' intentions relative to this norm on their perceptions of compartmentalization, a 2 (ingroup norm: prodownloading/no-downloading)  $\times$  2 (congruence: congruent/incongruent) ANOVA was conducted on the measure of compartmentalization. The main effects of the ingroup norm,  $F(1, 108) = 0.92$ ,  $p = .338$ ,  $\eta^2 = .008$ , and congruence,  $F(1, 108) = 0.01$ ,  $p = .948$ ,  $\eta^2 = .000$ ,

were not significant. However, a significant interaction was observed between ingroup norm and congruence,  $F(1, 108) = 16.03$ ,  $p < .001$ ,  $\eta^2 = .129$ . Interpretation of this interaction (see Figure 2) revealed that when the ingroup norm was a no-downloading one, participants whose behavioral intentions were congruent with this norm (those who did not engage in the illegal behavior) reported lower compartmentalization ( $M = 2.44$ ;  $SD = 1.25$ ) than participants whose behavioral intentions were incongruent with this norm (and who did engage in illegal down-

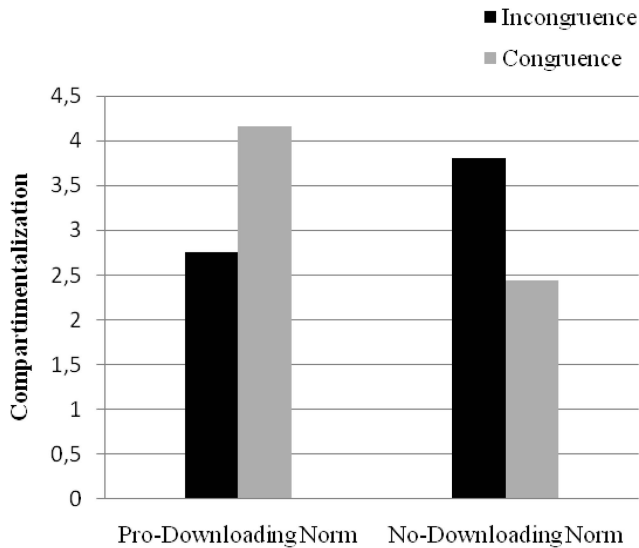


Figure 2. Compartmentalization as a function of ingroup norm and congruence.

loading;  $M = 3.81$ ;  $SD = 1.92$ ;  $F(1, 55) = 10.13$ ,  $p = .002$ ,  $\eta^2 = .156$ ). In contrast, when the ingroup norm was a prodownloading one, participants whose behavioral intentions were congruent with this norm (and who engaged in the illegal behavior) reported greater compartmentalization ( $M = 4.16$ ;  $SD = 1.62$ ) than participants whose behavioral intentions were incongruent with this norm (and who did not engage in illegal downloading;  $M = 2.75$ ;  $SD = 1.59$ ,  $F(1, 55) = 6.72$ ,  $p = .012$ ,  $\eta^2 = .112$ ). These findings support H3.

### Intraindividual Conflict

To assess the impact of the ingroup norm and the congruence of participants' intentions relative to this norm on their feelings of intraindividual conflict, a 2 (ingroup norm: prodownloading/no-downloading)  $\times$  2 (congruence: congruent/incongruent) ANOVA was conducted on the conflict variable. The main effects of the ingroup norm,  $F(1, 107) = .01$ ,  $p = .942$ ,  $\eta^2 = .000$ , and congruence,  $F(1, 107) = .31$ ,  $p = .577$ ,  $\eta^2 = .003$ , were not significant. However, a significant interaction was observed between ingroup norm and congruence,  $F(1, 107) = 15.07$ ,  $p < .001$ ,  $\eta^2 = .12$ . Interpretation of this interaction (see Figure 3) revealed that participants whose behavioral intentions were congruent with the no-downloading norm

(those who did not engage in the illegal behavior) reported lower conflict ( $M = 1.95$ ;  $SD = 1.04$ ) than participants whose behavioral intentions were incongruent with this norm (and who did engage in the illegal behavior;  $M = 3.18$ ;  $SD = 1.68$ ;  $F(1, 54) = 10.66$ ,  $p = .002$ ,  $\eta^2 = .165$ ). In contrast, when the ingroup norm was a prodownloading one, participants whose behavioral intentions were congruent with this norm (and who did engage in the illegal behavior) reported greater conflict ( $M = 3.01$ ;  $SD = 1.22$ ) than participants whose behavioral intentions were incongruent with this norm (and who did not engage in illegal behavior;  $M = 2.09$ ;  $SD = 0.67$ ,  $F(1, 53) = 5.64$ ,  $p = .021$ ,  $\eta^2 = .096$ ), supporting H4.

### Perceived Influence of Different Sources

To assess the impact of group norms and the congruence of participants' intentions relative to the ingroup norm on their perceptions of the influence of different sources, we conducted a 2 (ingroup norm: prodownloading/no-downloading)  $\times$  2 (congruence: congruent/incongruent) ANOVA with the five sources of normative influence as a repeated-measures factor. A significant main effect of the ingroup norm emerged,  $F(1, 104) = 5.02$ ,  $p = .027$ ,  $\eta^2 = .046$ , showing that participants generally perceived these sources of influence to be stronger

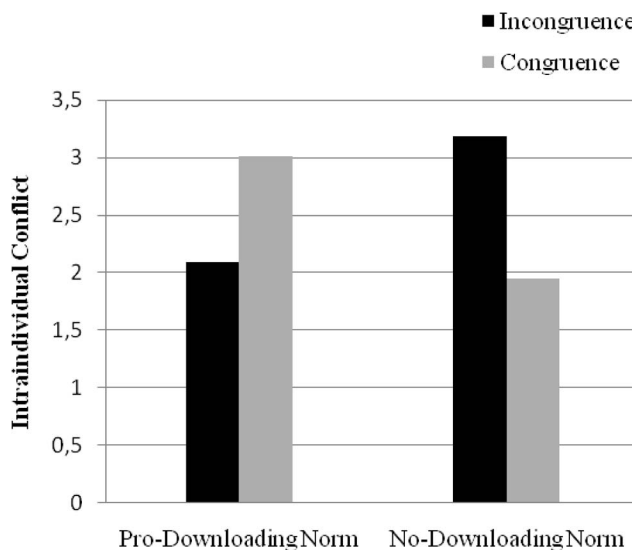


Figure 3. Intraindividual conflict as a function of ingroup norm and congruence.

when the ingroup norm was not supporting illegal downloading ( $M = 4.26$ ;  $SD = 0.17$ ) than when this norm was supporting illegal downloading ( $M = 3.77$ ;  $SD = 0.17$ ). A significant main effect for the source of influence emerged,  $F(5, 100) = 48.38$ ,  $p < .001$ ,  $\eta^2 = .708$ . Planned contrasts conducted to locate the differences among the sources showed that participants perceived the sources of influence pertaining to the situation ( $M = 5.04$ ;  $SD = 0.16$ ) and to their personal values ( $M = 4.97$ ;  $SD = 0.17$ ) to be stronger than their morality ( $M = 4.77$ ;  $SD = 0.19$ ;  $F(1, 10) = 3.49$ ,  $p = .029$ ,  $\eta^2 = .045$ ). Moreover, morality was seen as a more influential source than software companies ( $M = 3.19$ ;  $SD = 0.22$ ;  $F(1, 10) = 22.16$ ,  $p = .043$ ,  $\eta^2 = .039$ ). Finally, students from participants' university were perceived as having the least influence on their own behavior ( $M = 2.41$ ;  $SD = 0.18$ ;  $F(1, 104) = 52.34$ ,  $p = .006$ ,  $\eta^2 = .070$ ).

A three-way interaction emerged between ingroup norm, congruence, and source of normative influence,  $F(5, 100) = 4.36$ ,  $p = .001$ ,  $\eta^2 = .179$ . Simple effect tests revealed that for the sources of normative influence pertaining to software companies and to students, the norm

by congruence interactions were not significant,  $F_s = .75$  to  $1.34$ ,  $p_s = .25$  to  $.39$ ,  $\eta^2 = .007$  to  $.01$ . For the 3 other sources (i.e., personal values, morality, situation), significant norm by congruence interactions emerged (Figure 4;  $F_s = 10.117$  to  $13.28$ ,  $p_s = .000$  to  $.026$ ,  $\eta^2 = .04$  to  $.11$ ). Participants in the prodownloading norm condition who intended to behave congruently with this norm perceived that it was the situation ( $M = 5.46$ ;  $SD = 1.50$ ) that guided them more than their personal values ( $M = 3.93$ ;  $SD = 1.79$ ;  $F(1, 40) = 22.98$ ,  $p < .001$ ,  $\eta^2 = .37$ ), followed by their morality ( $M = 3.66$ ;  $SD = 1.93$ ;  $F(1, 40) = 4.91$ ,  $p = .032$ ,  $\eta^2 = .10$ ). In contrast, participants who intended to behave congruently with a no-downloading norm perceived that their personal values ( $M = 5.89$ ;  $SD = 1.45$ ) and their morality ( $M = 5.89$ ;  $SD = 1.44$ ;  $F(1, 27) = 00$ ,  $p = 1.00$ ,  $\eta^2 = .00$ ) guided them more in their behavior than the situation ( $M = 4.79$ ;  $SD = 1.55$ ;  $F(1, 27) = 10.45$ ,  $p = .003$ ,  $\eta^2 = .28$ ). There is no significant difference for participants who intended to behave either incongruently with a no-downloading or a prodownloading norm ( $F_s = 2.33$  to  $2.40$ ,  $p_s = .10$  to  $.17$ ,  $\eta^2 = .08$  to  $.13$ ).

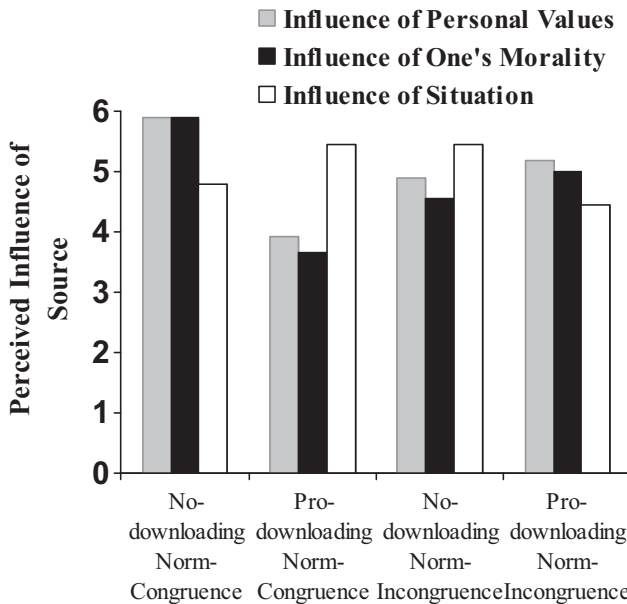


Figure 4. Perceived influence as a function of ingroup norm, congruence, and three sources of influence (personal values, morality, situation).

## Discussion

The first objective of this experiment was to investigate the motives that underlie an illegal normative behavior (i.e., illegally downloading computer software or not) and to measure the extent to which this behavior can be situationally internalized in the self and endorsed for self-determined motivations. As a second objective, this experiment tested the impact of a manipulated norm and of participants' intentions to behave in congruence with this norm on variables that captured participants' subjective perceptions in the internalization process, namely compartmentalization, intraindividual conflict, and perceived sources of influence. This study used an established methodology from the social norms literature (Smith & Louis, 2008). Our findings confirmed the role played by behavioral congruence in the internalization of a potentially socially harmful behavior, while providing new information on participants' subjective perceptions in this process. Investigating these ideas allows a better comprehension of the motivations that drive some illegal behaviors. In addition, identifying the subjective processes that are activated when people plan to act illegally allows to capture what the perpetrators of these actions experience when engaging in such behaviors, and to develop more targeted intervention programs aimed at reducing illegal behavior such as theft.

The ingroup norms we manipulated did influence participants' intentions to engage in prodownloading versus no-downloading behaviors, confirming H1. Also, participants who intended to behave congruently with their ingroup's no-downloading norm were more self-determined compared to participants whose behavioral intentions were not congruent with this norm (i.e., and who planned to behave illegally instead), which confirms H2a. In addition, participants whose behavioral intentions were incongruent with the prodownloading norm reported higher self-determination compared to participants whose behavioral intentions were congruent with this prodownloading ingroup norm, supporting H2b. Hence, above and beyond the normative pressure, individuals internalize more easily a prolegal than an illegal behavior. These results correspond with SDT and replicate Amiot and colleagues' findings (2012), by

showing that prosocial behaviors tend to be more self-determined, even if the individual is acting against his or her group norm. Moreover, these prior studies directly measured the behavior, whereas the current study assessed behavioral intentions. It is reassuring to observe similar results across the studies.

The current results are also consistent with literature that states that an individual may deviate from his or her group norms if s/he considers that it is better for the group to do so; such as because doing so may improve the group situation (Packer, 2008), or if the ingroup norms go against one's own moral values (Hornsey, Smith, & Begg, 2007). Still, it would be interesting to further investigate whether the norm that encourage illegally downloading computer software is perceived as being acceptable for the individual and as being hurtful versus good for the group. Even if it is illegal, accessing the latest software can be perceived as good when it help the group to maintain high standards which may then allow its members to gain social recognition and status. The perception of how the norm helps (vs. hurts) the group may influence the internalization process by facilitating the acceptance or rejection of the norm within his or her self. Furthermore, deviating from ingroup norms when these norms are perceived as harmful may represent a strategy that group members use to realign their group toward more prosocial actions and norms (Packer, 2008).

Moreover, participants who intended to act illegally—both by behaving incongruently with a no-downloading norm and by failing to follow a no-downloading norm—felt more compartmentalization and intraindividual conflict than participants who did not intend to engage in the illegal downloading behavior (supporting H3 and H4). These results confirm that compartmentalization and intraindividual conflict are most strongly activated when participants intend to behave illegally, regardless of ingroup norms: Behaving incongruently with the norm of an important ingroup per se did not induce such high levels of compartmentalization or conflict. These variables may help to protect the self from the negative impact of behaving illegally. While the current data do not provide exact information about which specific mechanisms may be blocking the internalization of illegal norms and also making these norms

harder to feel comfortable about (e.g., intuitive knowledge about what's right vs. wrong), this is an exciting direction for future research which align with SDT postulates.

Participants' perceptions of the five sources of influence also differed as a function of norms and congruence. When participants were intending to behave legally, they were more likely to perceive that it was their personal values/morality that guided their behavior. When they intended to behave illegally, they stated that the situational sources influenced them the most. These findings could be attributable to a classic self-serving bias (Miller & Ross, 1975), whereby people attribute their positive behaviors to internal causes (their values) and attribute their negative behaviors to external causes (the situation). These findings could also relate to the notion of compartmentalization; when participants behave illegally, they may become more prone to perceive the situation as influencing them and to dissociate themselves personally from their action by putting a distance between their self (i.e., their own values and morality) and their behavior.

Interestingly, and despite the fact that our ingroup norm manipulation had a strong effect on participants' intentions in the experiment, participants were less likely to perceive that students from their university had influenced their behavioral intentions. This pattern of findings is also congruent with literature suggesting that explicit and implicit processes are not always congruent (Briñol, Petty, & Wheeler, 2006) and that what is really influencing a person's behavior is often different from what the person's subjectively reported as influencing him/her (Nisbett & Wilson, 1977). Still, the correlation between identifying with the student ingroup and perceiving the group as a source of influence suggest that the more participants identify with this ingroup, the more they consider the student group as influencing them, which is coherent with social identity theory and its approach to group norms (e.g., Terry & Hogg, 1996; Turner, 1991). Future studies should further investigate these subjective perceptions, how they shape and predict different types of normative behaviors, and what are the potential effects of this mismatch between perceptions and behaviors.

It will also be interesting to further test whether the results obtained herein can be ob-

served in different life contexts and over the longer-term. It is important to note that the current research investigated the immediate impact of a situational norm in an experimental context. Still, we can wonder about the process of internalization over time in real life contexts and of how repeating behaviors over time may lead to the internalization of these behaviors. Further studies should investigate these longer-term processes by using longitudinal design. For example, we could follow over a 1-year period a cohort of new university students as they are in the process of endorsing new social norms, including potentially harmful norms (e.g., cheating in exams). By assessing repeatedly over this period of time variables such as perceived norms, the frequency of the normative behavior and self-determination to engage in this behavior, and degree of social identification with the group, such research will allow to paint a more complete picture of the internalization process over time. Moreover, the rate of internalization of group norms is possibly linked to changes in identification with the group, such that a normative behavior becomes more congruent and internalized in the person's self when this person develops a stronger sense of identification with the group (Terry & Hogg, 1996). Understanding the temporal and longer-term processes involved in norm internalization could be useful to eventually create programs or campaigns that aim to reduce illegal behaviors in the workplace or school context. Considering the immediate versus the long-term impact of a group norm on the behaviors and their internalization will also provide a more complete account for why individuals engage in those negative behaviors and how such behaviors develop.

## References

- Aguiar, L., & Martens, B. (2013). *Digital music consumption on the Internet: Evidence from Clickstream Data*. European Commission: Institute for prospective technological studies. doi:10.2791/83798
- Ajzen, I. (2001). Nature and operation of attitudes. *Annual Review of Psychology*, 52, 27–58.
- Amiot, C. E., & Bourhis, R. Y. (2003). Discrimination and the positive-negative asymmetry effect: Ideological and normative processes. *Personality and Social Psychology Bulletin*, 29, 597–608. doi:10.1177/0146167203029005005

- Amiot, C. E., & Sansfaçon, S. (2011). Motivations to identify with social groups: Positive and negative consequences. *Group Dynamics: Theory, Research, and Practice*, 15, 105–127. doi:10.1037/a0023158
- Amiot, C. E., Sansfaçon, S., & Louis, W. R. (2013). Investigating the motivations underlying harmful social behaviors and the motivational nature of social norms. *Journal of Applied Social Psychology*, 43, 2146–2157. doi:10.1111/jasp.12167
- Amiot, C. E., Sansfaçon, S., & Louis, W. R. (in press). How normative and social identification processes predict self-determination to engage in derogatory behaviours against outgroup hockey fans. *European Journal of Social Psychology*.
- Amiot, C. E., Sansfaçon, S., Louis, W. R., & Yelle, M. (2012). Testing the role of group norms and behavioral congruence in the internalization of discriminatory and parity behaviors. *Personality and Social Psychology Bulletin*, 38, 63–76. doi:10.1177/0146167211429804
- Baer, J. S., Stacy, A., & Larimer, M. (1991). Biases in the perception of drinking norms among college students. *Journal of Studies on Alcohol*, 52, 580–586.
- Baumeister, R. F., & Vohs, K. D. (2004). Four roots of evil. In M. G. Arthur (Ed.), *The social psychology of good and evil*, (pp. 87–101). New York, NY: Guilford Press.
- Belgian Anti-Piracy Federation. (2009). *Étude sur le téléchargement auprès des jeunes en Belgique (Study on illegal downloading in Belgium Youth)*. Press Release.
- Benet-Martínez, V., & Haritatos, J. (2005). Bicultural Identity Integration (BII): Components and socio-personality antecedents. *Journal of Personality*, 73, 1015–1050. doi:10.1111/j.1467-6494.2005.00337.x
- Briñol, P., Petty, R. E., & Wheeler, S. C. (2006). Discrepancies between explicit and implicit self concepts: Consequences for information processing. *Journal of Personality and Social Psychology*, 91, 154–170. doi:10.1037/0022-3514.91.1.154
- Copyright Board of Canada. (2009). *Copying for Private Use*. Copyright Art of Canada, subsection, 83(8).
- Deci, E. L., & Ryan, R. M. (2000). The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11, 227–268. doi:10.1207/S15327965PLI1104\_01
- Fielding, K. S., McDonald, R., & Louis, W. R. (2008). Theory of planned behavior, identity, and intentions to engage in environmental activism. *Journal of Environmental Psychology*, 28, 318–326. doi:10.1016/j.jenvp.2008.03.003
- Frone, M. R., & Brown, A. L. (2010). Workplace substance-use norms as predictors of employee substance use and impairment: A survey of US workers. *Journal of Studies on Alcohol and Drugs*, 71, 526–534.
- Guay, F., Mageau, G. A., & Vallerand, R. J. (2003). On the hierarchical structure of self determined motivation: A test of top-down, bottom-up, reciprocal, and horizontal effects. *Personality and Social Psychology Bulletin*, 29, 992–1004. doi:10.1177/0146167203253297
- Haslam, S. A., Oakes, P. J., Reynolds, K. J., & Turner, J. C. (1999). Social identity salience and the emergence of stereotype consensus. *Personality and Social Psychology Bulletin*, 25, 809–818. doi:10.1177/0146167299025007004
- Haslam, S. A., & Reicher, S. (2007). Beyond the banality of evil: Three dynamics of an integrationist social psychology of tyranny. *Personality and Social Psychology Bulletin*, 33, 615–622. doi:10.1177/0146167206298570
- Hornsey, M. J., Smith, J. R., & Begg, D. (2007). Effects of norms among those with moral conviction: Counter-conformity emerges on intentions but not behaviors. *Social Influence*, 2, 244–268. doi:10.1080/15534510701476500
- Jacobson, R. J., Mortensen, C. R., & Cialdini, R. (2011). Bodies obliged and unbound: Differentiated response tendencies for injunctive and descriptive social norms. *Journal of Personality and Social Psychology*, 100, 433–448.
- JDN. (2004). *Téléchargement illégal: On continue malgré les risques*. (Illegal downloading: We continue regardless of the risks.) Retrieved on September 20, 2013, from <http://www.journaldunet.com/0405/040514sondage.shtml>
- Kelman, H. C. (1961). Processes of opinion change. *Public Opinion Quarterly*, 25, 57–78. doi:10.1086/266996
- Kerr, N. L., Garst, J., Kiehle, D., & Harris, S. (1997). That still, small voice: Commitment to cooperate as an internalized vs. a social norm. *Personality and Social Psychology Bulletin*, 23, 1300–1311. doi:10.1177/01461672972312007
- Kerr, N. L., & Kaufman-Gilliland, C. M. (1994). Communication, commitment, and cooperation in social dilemmas. *Journal of Personality and Social Psychology*, 66, 513–529. doi:10.1037/0022-3514.66.3.513
- Kura, K. M., Shamsudin, F. M., & Chauhan, A. (2013). Modeling the influence of group norms and self-regulatory efficacy on workplace deviant behaviour. *Asian Social Science*, 9, 113–122.
- Lee, K., & Allen, N. J. (2002). Organizational citizenship behavior and workplace deviance: The role of affect and cognition. *Journal of Applied Psychology*, 87, 131–142. doi:10.1037/0021-9010.87.1.131
- Lian, H., Ferris, L. D., & Brown, D. J. (2012). Does taking the good with the bad make things worse? How abusive supervision and leader-member ex-



- change interact to impact need satisfaction and organizational deviance. *Organizational Behavior and Human Decision Processes*, 117, 41–52. doi:10.1016/j.obhdp.2011.10.003
- Lynch, M. F., La Guardia, J. G., & Ryan, R. M. (2009). On being yourself in different cultures: Ideal and actual self-concept, autonomy support, and well-being in China, Russia, and the United States. *The Journal of Positive Psychology*, 4, 290–304. doi:10.1080/17439760902933765
- Miller, D. T., & Ross, M. (1975). Sell-serving biases in the attribution of causality: Fact or fiction? *Psychological Bulletin*, 82, 213–225. doi:10.1037/h0076486
- Millette, V., & Gagné, M. (2008). Designing volunteers' tasks to maximize motivation, satisfaction and performance: The impact of job characteristics on volunteer engagement. *Motivation and Emotion*, 32, 11–22. doi:10.1007/s11031-007-9079-4
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 1–259.
- Packer, D. J. (2008). On being both with us and against us: A normative conflict model of dissent in social groups. *Personality and Social Psychology Review*, 12, 50–72. doi:10.1177/1088868307309606
- Reno, R. R., Cialdini, R. B., & Kallgren, C. A. (1993). The transsituational influence of social norms. *Journal of Personality and Social Psychology*, 64, 104–112. doi:10.1037/0022-3514.64.1.104
- Robinson, S. L., & Greenberg, J. (1998). Employees behaving badly: Dimensions, determinants and dilemmas in the study of workplace deviance. In C. L. Cooper & D. M. Rousseau (Eds.), *Trends in organizational behavior* (Vol. 5, pp. 1–30). New York, NY: Wiley.
- Robinson, S. L., & O'Leary-Kelly, A. M. (1998). Monkey see, monkey do: The influence of work groups on the antisocial behavior of employees. *Academy of Management Journal*, 41, 658–672. doi:10.2307/256963
- Ryan, R. M., & Connell, J. P. (1989). Perceived locus of causality and internalization: Examining reasons for acting in two domains. *Journal of Personality and Social Psychology*, 57, 749–761. doi:10.1037/0022-3514.57.5.749
- Ryan, R. M., & Deci, E. L. (2003). On assimilating identities to the self: A self-determination theory perspective on internalization and integrity within cultures. In M. R. Ryan & E. L. Deci (Eds.), *Handbook on self & identity* (pp. 253–272). New York, NY: The Guilford Press.
- Smith, J. R., Hogg, M. A., Martin, R., & Terry, D. J. (2007). Uncertainty and the influence of group norms in the attitude-behaviour relationship. *British Journal of Social Psychology*, 46, 769–792. doi:10.1348/014466606X164439
- Smith, J. R., & Louis, W. R. (2008). Do as we say and as we do: The interplay of descriptive and injunctive group norms in the attitude-behavior relationship. *British Journal of Social Psychology*, 47, 647–666. doi:10.1348/014466607X269748
- Smith, J. R., Terry, D. J., Manstead, A. S. R., Louis, W. R., Kotterman, D., & Wolfs, J. (2008). The attitude-behavior relationship in consumer conduct: The role of norms, past behavior, and self-identity. *The Journal of Social Psychology*, 148, 311–333. doi:10.3200/SOCP.148.3.311-334
- Spears, R., Postmes, T., Lea, M., & Wolbert, A. (2002). When are net effects of gross products? The power of influence and the influence of power in computer-mediated communication. *Journal of Social Issues*, 58(1), 91–107. doi:10.1111/1540-4560.00250
- Tajfel, H., & Turner, J. C. (1986). The social identity theory of inter-group behavior. In S. Worchel & L. W. Austin (Eds.), *Psychology of intergroup relations*. Chicago, IL: Nelson-Hall.
- Terry, D. J., & Hogg, M. A. (1996). Group norms and the attitude-behavior relationship: A role for group identification. *Personality and Social Psychology Bulletin*, 22, 776–793. doi:10.1177/0146167296228002
- Tripp, B. (2009). Fathers in jail: Managing dual identities. *Applied Psychology in Criminal Justice*, 5, 26–56.
- Turner, J. C. (1991). *Social influence*. Belmont, CA: Thomson Brooks/Cole Publishing Co.
- Vallerand, R. J. (1997). Toward a hierarchical model of intrinsic and extrinsic motivation. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 29, pp. 271–360). San Diego, CA: Academic Press. doi:10.1016/S0065-2601(08)60019-2
- Vallerand, R. J., & Bissonnette, R. (1992). Intrinsic, extrinsic, and amotivational styles as predictors of behavior: A prospective study. *Journal of Personality*, 60, 599–620. doi:10.1111/j.1467-6494.1992.tb00922.x
- Van Bavel, J. J., Packer, D. J., & Cunningham, W. A. (2008). The neural substrates of in-group bias: A functional magnetic resonance imaging investigation. *Psychological Science*, 19, 1131–1139. doi:10.1111/j.1467-9280.2008.02214.x
- Vorauer, J. D., & Miller, D. T. (1997). Failure to recognize the effect of implicit social influence on the presentation of self. *Journal of Personality and Social Psychology*, 73, 281–295. doi:10.1037/0022-3514.73.2.281

Received September 24, 2012

Revision received December 14, 2013

Accepted December 17, 2013 ■