

2012

Good and Bad Group Performance: Same Process—Different Outcomes

R. S. Tindale

Loyola University Chicago, rtindal@luc.edu

Christine M. Smith

Grand Valley State University, smithc@gvsu.edu

Amanda Dykema-Engblade

Northeastern Illinois University

Katharina Kluwe

Loyola University Chicago

Follow this and additional works at: https://scholarworks.gvsu.edu/psy_articles



Part of the [Psychology Commons](#)

Recommended Citation

Tindale, R. S.; Smith, Christine M.; Dykema-Engblade, Amanda; and Kluwe, Katharina, "Good and Bad Group Performance: Same Process—Different Outcomes" (2012). *Peer Reviewed Articles*. 37.

https://scholarworks.gvsu.edu/psy_articles/37

This Article is brought to you for free and open access by the Psychology Department at ScholarWorks@GVSU. It has been accepted for inclusion in Peer Reviewed Articles by an authorized administrator of ScholarWorks@GVSU. For more information, please contact scholarworks@gvsu.edu.

Good and bad group performance: Same process—different outcomes

Group Processes & Intergroup Relations

15(5) 603–618

© The Author(s) 2012

Reprints and permission:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/1368430212454928

gpir.sagepub.com



R. Scott Tindale,¹ Christine M. Smith,² Amanda Dykema-Engblade³ and Katharina Kluwe¹

Abstract

Much of the research on small group performance shows that groups tend to outperform individuals in most task domains. However, there is also evidence that groups sometimes perform worse than individuals, occasionally with severe negative consequences. Theoretical attempts to explain such negative performance events have tended to point to characteristics of the group or the group process that were different than those found for better performing groups. We argue that typical group processes can be used to explain both good and bad group performance in many instances. Results from a pair of experiments focusing on two different task domains are reported and used to support our arguments.

Keywords

group performance

Paper received 21 June 2012; revised version accepted 21 June 2012.

Groups are a ubiquitous aspect of human existence (Brewer & Caporael, 2006). We are born and raised in families, learn in classrooms, socialize in friendship groups and communities, and often work in teams. One of the reasons that work is done in teams is that the resources available to a team are superior to those harnessed by a single individual (Wegner, 1987). These additional resources allow group performance to exceed that of a single individual in most task domains (Davis, 1969; Larson, 2010). The vast amount of research on group performance shows that groups, although rarely performing at their full potential, typically perform as well or better than individuals on most tasks (Hill, 1982; Hinsz, Tindale, &

Vollrath, 1997; Kerr & Tindale, 2004). Thus, groups are used in many key aspects of both public (e.g., parliaments, cabinets, juries, etc.) and private (e.g., task forces, corporate boards, focus groups, etc.) life.

¹Loyola University Chicago, USA

²Grand Valley State University, USA

³Northeastern Illinois University, USA

Corresponding author:

R. Scott Tindale, Department of Psychology, Loyola University Chicago, 1032 W. Sheridan Rd., Chicago, IL 60660 USA.

Email: rtindal@luc.edu

Groups do not always perform well, however. There are many tasks where groups are inappropriate, or at least extremely inefficient (e.g., writing a poem). And without appropriate resources, even well designed teams may fail (Hackman, 1998). But even well-resourced groups with knowledgeable members sometimes will make poor (even disastrous) decisions (Janis, 1982). Incidents like the space shuttles *Challenger* and *Columbia* explosions and the failed *Bay of Pigs* invasion are examples of poor group decisions. In each case, groups of experts had made or supported the decisions even though, allegedly, information was present that should have led them to decide otherwise (Nijstad, 2009). Such decisions are often blamed on poor decision processes. Probably the most well-known description of dysfunctional group decision processes is Janis' (1982) *groupthink*. Based on a number of case studies, Janis defined multiple aspects of groupthink including high cohesiveness, directive leadership, poor information search, and sanctions against dissent. Although there is now evidence that poor information search and sanctions against dissent can impede group performance under some circumstances (see Brodbeck, Kerschreiter, Mojzisch, Frey, & Schulz-Hardt, 2007; Postmes, Spears, & Cihangir, 2001), neither insures that groups will do poorly. In addition, most evidence argues that high cohesiveness and directive leadership often aid group performance (Mullen & Copper, 1994; Perterson, 1997). Thus, the research evidence to date has not been particularly supportive of groupthink either as a phenomenon or an explanation of poor group performance (Baron, 2005; Kerr & Tindale, 2004).

Recent theorizing has begun to explore how basic group processes can be used to explain both good and poor group performance in different contexts (Brodbeck et al., 2007; Kerr and Tindale, 2004; Tindale, Talbot, & Martinez, in press). In the current paper, we argue that two rather basic and quite common aspects of group consensus processes can be used to explain many instances of both good and poor group performance. Both aspects can be seen as instances of what we have referred to as "social sharedness" (Kameda,

Tindale, & Davis, 2003; Tindale & Kameda, 2000). Social sharedness is the idea that task-relevant cognitions (broadly defined) that the members of a group have in common, or share, exert a greater influence on the group than do similar cognitions that are not shared among the members. The cognitions that are shared can vary from preferences for decision alternatives or information about the alternatives to heuristic information processing strategies that the members cannot even articulate. However, the greater the degree of sharedness for a particular task relevant cognition, the greater the likelihood that it will influence the group decision. In general, we will argue that social sharedness is often adaptive and probably evolved as a useful aspect of living in groups (Kameda & Tindale, 2006; Kameda, Wisdom, Toyokawa, Inukai, 2012). However, when the shared cognition is inappropriate to the current situation, it can lead groups to make poor decisions.

The current research focusses on two types or levels of social sharedness: Shared preferences and shared task representations (Tindale, Smith, Steiner, Filkins, & Sheffey, 1996). Shared preferences refer to the degree to which members of a group prefer a particular decision alternative. Numerous studies have shown that the size of a faction favoring a particular alternative is a good predictor of the likelihood that the group will choose that alternative, and that the largest faction defines the group consensus most of the time (Davis, Kerr, Atkin, Holt, & Meek, 1975; Tindale, Davis, Vollrath, Nagao, & Hinsz, 1990). Thus, a majority/plurality decision model (or social decision scheme – SDS, Davis, 1973) does relatively well in predicting group decision outcomes on a variety of tasks. Table 1 shows the SDS representation (the social decision scheme matrix, **D**; Davis, 1973) for a majority wins process for a six-person group. It also shows a proportionality SDS that represents the probability of group decision outcomes as identical to the proportion of members that initially favored the alternative. A proportionality model provides a good baseline for studies comparing individual and group decision making since it predicts group

Table 1. Social Decision Scheme Models for Majority Wins–Otherwise Equiprobability and Proportionality

Individual distributions A–B	Group distributions			
	Majority Wins – Equiprobability Otherwise		Proportionality	
	A	B	A	B
6–0	1.00	0.00	1.00	0.00
5–1	1.00	0.00	0.87	.013
4–2	1.00	0.00	0.67	.033
3–3	0.50	0.50	0.50	0.50
2–4	0.00	1.00	0.33	0.67
1–5	0.00	1.00	0.13	0.87
0–6	0.00	1.00	0.00	1.00

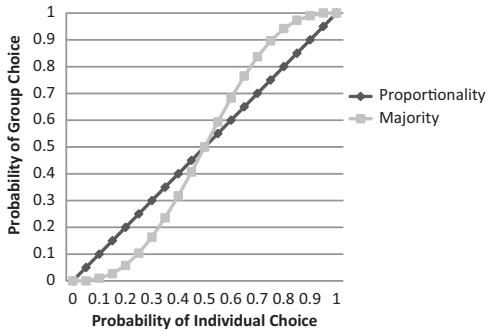


Figure 1. The probability of the group choosing A as a function of individual choice probabilities for both Majority Wins–Otherwise Equiprobability and Proportionality Social Decision Schemes.

and individual preference distributions to be equivalent. As can be seen in the table, any time four or more individual favor a given alternative (A or B), a majority model predicts the group will choose that alternative.

Figure 1 shows the predicted probabilities of the group choosing alternative A for both the majority and the proportionality models as a function of the probability that a randomly sampled individual from the population from which group members would be drawn would choose A. Given the proportionality model functions as the individual choice baseline, the figure shows that when individuals are likely to choose A, a group employing a majority-wins process

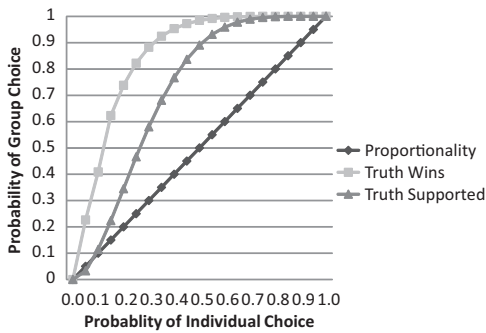
exacerbates this tendency and is even more likely to choose A. Under the assumption that A is the correct or optimal alternative, majority-wins groups should perform better than the average individual anytime the individual choice probability for A is greater than .5. However, if individual preference probabilities favor B, then majority-wins groups will choose A less often than individuals, thus performing more poorly.

The second aspect of social sharedness relevant to the current research involves what we have called “shared task representations” (Tindale et al., 1996). Tindale and colleagues defined a shared task representation as “any task/situation relevant concept, norm, perspective, or cognitive process that is shared by most or all of the group members (Tindale et al., 1996, p. 84). “Task/situation relevant” means that the representation must have implications for the choice alternatives involved, and the degree to which a shared representation affects group decision processes and outcomes will vary as a function of its perceived relevance. Its influence will also vary by the degree to which it is shared among the group members—the greater the degree of sharedness (the more members who share it), the greater its influence.

Probably the best example of shared task representation is the first component of Laughlin and Ellis’ (1986) definition of a demonstrable task—a task where group members can demonstrate during group discussion that a particular alternative is “correct” or “optimal”. Laughlin (1999) has argued that one of the reasons that groups are better problem solvers than are individuals is that group members often share a conceptual system (i.e., a shared task representation) that allows them to realize when a proposed solution is correct within that system. This shared conceptual system, or background knowledge, is what allows a minority member with a correct answer to influence a larger incorrect faction to change its preference to the correct alternative. Such situations are well described by SDS models called “truth wins” and “truth supported wins” (Laughlin, 1980; see Table 2). Truth wins predicts that any group that has at least one member with

Table 2. Social Decision Scheme Models for Truth Wins and Truth-Supported Wins

Individual distributions	Group distributions			
	Truth Wins		Truth-Supported Wins	
	A	B	A	B
A–B	A	B	A	B
6–0	1.00	0.00	1.00	0.00
5–1	1.00	0.00	1.00	0.00
4–2	1.00	0.00	1.00	0.00
3–3	1.00	0.00	1.00	0.00
2–4	1.00	0.00	1.00	0.00
1–5	1.00	0.00	0.00	1.00
0–6	0.00	1.00	0.00	1.00

**Figure 2.** The probability of the group choosing the correct alternative as a function of individual choice probabilities for Truth Wins, Truth-Supported Wins and Proportionality Social Decision Schemes.

the correct answer will be able to solve the problem correctly (Lorge & Solomon, 1955; Laughlin, 1980). Truth supported wins argues that at least two members of the group must have the correct answer in order for the group to solve the problem correctly (Laughlin, 1980). Figure 2 presents the relationship between individual and six-person group decision preferences (for alternative A which is assumed correct in the present context) under both a truth wins and truth-supported wins SDS. In both cases, the group probability of choosing the correct response increases more rapidly than the probability of an individual choosing the correct response. The probability increase for groups under the truth wins model is considerably steeper than the

increase for the truth supported wins model, but both curves show that groups will virtually always out-perform individuals when correct minority faction can convince incorrect majority factions to switch their preferences to the correct alternative. However, if one were to assume that the shared task representation favored an incorrect, rather than a correct, alternative (assume alternative A is incorrect), then Figure 2 would represent an “error wins” process and groups would rarely if ever outperform individuals.

Research has shown that shared task representations do not always favor normatively or objectively correct or optimal alternatives (Hinsz, Tindale, & Nagao, 2008; Tindale, 1993). Using problems where individuals tend to make intuitively appealing but incorrect judgments, groups often show biases in the same intuitive but incorrect direction. Hinsz et al. (2008) asked individuals and groups to respond to base rate neglect problems similar to the “cab problem” employed by Tversky and Kahneman, (1980). In this problem, a witness to a hit and run accident says the cab involved was green, but there are only two colors of cabs in the city and there are many more blue cabs (85%) than green cabs (15%) on the road at any given point in time. In a later test, the witness is found to be 80% accurate in distinguishing between blue and green cabs. With less than perfect witness accuracy, estimates that the cab was actually green should be well below 80% yet many individuals anchor their estimates at or near 80%. Hinsz et al. found that groups were even less likely than individuals to temper their estimates in the direction of the base rate and were more likely to simply choose 80%. Smith, Tindale, and Steiner (1998) had individuals and five-person groups make investment decisions for “sunk cost” problems (problems where people tend to continue investing in a project that is failing because they have already invested a large amount of money—i.e., a “sunk cost”) and found that groups were equally likely as individuals to choose to put “good money after bad”. In addition, two-person minorities favoring continued investment were often persuasive against 3 person majorities favoring the opposite when they used sunk cost

arguments (e.g., “if we stop now all that money went to waste”).

Much of research on good vs. poor group performance has tended to argue that the different performance outcomes are a function of inherently different processes (Brodbeck et al., 2007; Janis, 1982). However, based on the work on shared preferences and shared representations, it is quite possible that many instances of poor performance are simply basic group processes working in a context where they lead the group in a poor direction. The two poor performance situations of interest here would be those where larger factions tended to be wrong and the shared representation would favor the incorrect alternative. In other words, the same basic processes that groups often use that lead to good performance outcomes can lead to poor outcomes in certain contexts. The two studies we report here both demonstrate this basic idea. Groups making decisions using shared representations of the task and how it should be approached, in conjunction with majority influence processes will lead to good outcomes when the representations aid in reaching good decisions, but will lead to poor outcomes when they do not. Both good and poor performance by groups (i.e., better and worse than an individual performance baseline) will be demonstrated using two different task domains; syllogistic reasoning and probability estimation.

Study 1

Study 1 is a follow up to a study comparing individual and group performance on syllogistic reasoning tasks across cultures where the premises were consistent or inconsistent with different cultural beliefs (Smith et al., 2000). In the Smith et al. study, individuals and groups worked on 8 syllogisms and were asked to judge whether the conclusion was valid or invalid. Half of the syllogisms had valid conclusions and the other half had invalid conclusions, based on formal, propositional logic. Generally, groups performed better than individuals and Japanese students performed better than their American counterparts. However,

there was one syllogism where Japanese groups performed substantially worse than individuals. This anomaly led us look more closely at that particular syllogism and we found that it was the only one we used that included the qualifier “some” in one of the premises and in the conclusion (e.g., some A are B). In formal logic, saying “some A are B” merely insures that the categories are not mutually exclusive, but does not exclude other possibilities (e.g., that “all A are B” or “All B are A”). However, research has shown that people often assume via conversational norms (Grice, 1975) that the inclusion of “some” in the statement implies more—e.g., that only some and not all are included (Evans, Newstead, & Byrne, 1993). In further analysing the group discussions, it appeared that such misinterpretations of the formal logic definition of “some A are B” led groups to perceive the syllogism’s conclusion as valid when it was not.

Misconstruing how the term “some” should be interpreted in formal logic should not always impede performance on syllogism problems. Syllogisms that use some in both a premise and the conclusion can be valid, and would be seen as valid even if “some” were misinterpreted. Thus, misconstruing the implications of “some” may actually aid in the correct evaluation of a syllogism under some circumstances. The current study attempted to assess whether this shared conversational norm would both aid and impair groups’ abilities to correctly evaluate the validity of different syllogisms.

Method

Participants Participants were 522 undergraduate introductory psychology students drawn from two Midwestern universities. Students participated either as individuals (132) or as members of five-person groups. All participants were given course credit for their participation.

Materials and procedures Upon arrival at the lab, participants were randomly assigned to either the individual or group condition. In both conditions, participants first responded to eight

sylogistic reasoning problems as individuals. All eight problems included the qualifier “some” in one of the premises and in the conclusion. Half of the syllogism contained a valid conclusion that logically follows from the premises, for example:

All protesters are healthy for the country.

Some radical leftists are protestors.

Therefore, some radical leftists are healthy for the country.

Regardless of whether the interpretation of the minor (second) premise implies “some are not”, the conclusion is still valid. The other half of the syllogisms included a conclusion that did not logically follow from the premises, but appeared valid under certain logically-erroneous but conversationally-plausible interpretations of what “some” means. For example:

Some immoral people are scientists.

All scientists are atheists.

Therefore, some atheists are not immoral.

For these syllogisms, erroneously assuming that “Some immoral people are scientists” implies that the two categories “immoral” and “scientists” overlap only partially (and, hence, that one could further assume that “Some immoral people are not scientists” or that “Some not immoral people are scientists”) leads to the appearance that the conclusion is valid. Half of the participants received the syllogisms in a pre-specified random order and the other half received them in the reverse order. Participants responded to each syllogism’s conclusion using an eight-point scale ranging from 1 (absolutely invalid) to 8 (absolutely valid) with scale scores 3 and 4 labeled “perhaps invalid” and 5 and 6 labeled “perhaps valid”.

Participants in the group condition worked on the eight syllogisms as a group and chose a score between 1 and 8 for each syllogism. Groups were told that they could reach their group decision in

any way that they liked but that the final group response should reflect the groups’ collective opinion. Participants in the individual condition worked on an unrelated task during the times while groups worked on the problems. Following the group responses or the work on the unrelated task, all participants again rated the eight syllogisms as individuals using the same response scale. Participants were told that these responses did not need to correspond to those they or their group had made before, but should reflect their current thinking on the problem. Following this final round, all participants were debriefed, thanked for their participation and excused from the experiment.

Results

For the four valid syllogisms, higher ratings were associated with better performance, while the opposite was true for the four invalid syllogisms. Thus, we reverse scored the invalid syllogisms so that higher ratings always defined better performance. We then averaged the ratings for both individuals and groups across the four syllogisms of each type. We then performed a 2 (group vs. 2nd test individuals) by 2 (syllogism type: valid vs. invalid) analysis of variance on the average scores. The means associated with condition are presented in Figure 3. The results showed a significant main effect for the individual—group difference, $F(1, 208) = 10.73, p < .01, \text{partial } \eta^2 = .049$, with groups ($M = 4.86$) performing better than individuals ($M = 4.57$). A significant main effect of syllogism type was also found, $F(1, 208) = 19.03, p < .001, \text{partial } \eta^2 = .084$, showing better performance for valid ($M = 5.01$) than invalid ($M = 4.35$) syllogisms. However, both main effects were qualified by a significant individual-group by type of syllogism interaction, $F(1, 208) = 16.36, p < .001, \text{partial } \eta^2 = .073$. As shown in Figure 3 and collaborated with simple effects analyses, groups performed better than individuals for valid syllogisms, $F(1, 208) = 22.08, p < .001, \text{partial } \eta^2 = .096$, yet performed worse than individuals for invalid syllogisms, $F(1, 208) = 6.46, p < .02, \text{partial } \eta^2 = .030$.¹

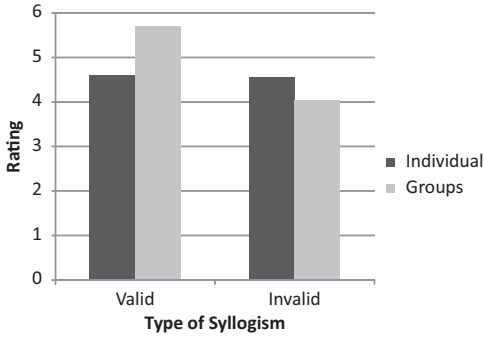


Figure 3. Individual and group performance for valid and invalid syllogisms

Table 3. Observed Social Decision Schemes for valid and invalid syllogisms

Individual distributions	Group distributions					
	Valid			Invalid		
	N	C	I	N	C	I
C-I						
5-0	38	1.00	0.00	25	0.72	0.28
4-1	92	0.96	0.04	62	0.73	0.27
3-2	64	0.81	0.19	79	0.35	0.65
2-3	72	0.65	0.35	77	0.42	0.58
1-4	34	0.41	0.59	50	0.24	0.76
0-5	3	0.00	1.00	11	0.09	0.91

C = Correct Response; I = Incorrect Response; N = Number of groups.

Two different approaches were used to assess aspects of the group discussion and influence processes. First, Table 3 shows the observed social decision scheme matrices for both the valid and invalid syllogisms. The initial preference distributions are based on group members' initial responses scored as correct vs. incorrect as a function of whether their judgments were on the "correct" half of the scale (1-4 for invalid syllogisms and 5-8 for valid syllogisms). Evidence for both faction size/majority processes and the effects of shared task representations are present. First, there were a greater number of groups with majorities favoring the correct response for the valid (64%) as opposed to the invalid (55%) syllogisms. In addition, the correct majorities won

93% of the time for the valid syllogisms, but only 55% of the time for the invalid syllogisms. A key comparison between the different types of syllogisms can be seen for the groups where a correct three-person majority faced an incorrect two person minority. For the valid syllogism, the majority prevailed 81% for the time, but for the invalid syllogisms, the majority won only 35% of the time, $\chi^2(1, N = 143) = 33.55, p < .001, \Phi = .48$. Such differences tend to imply that the incorrect response was easier to defend for the invalid syllogisms.

The group discussions were also videotaped and coded for whether the groups discussed the meaning of "some" in the context of the syllogisms (Cohen's $\kappa = .75$).² In all cases where some was mentioned, it was resolved consistent with conversational norms rather than formal logic (i.e., they always agreed that some implies some are not). For the valid syllogisms, when some was discussed, it improved group performance (93% correct) relative to when it was not discussed (64%), $\chi^2(1, N = 42) = 3.94, p < .05, \Phi = .31$. For invalid syllogisms, this trend was reversed, though not significantly so, with groups that mentioned some (14%) doing less well than group that did not mention some (32%), $\chi^2(42) = 1.54, p > .05, \Phi = .19$.

The results generally showed that within the same problem domain, groups could perform both better and worse than individuals depending on whether their shared understanding of the problems tended to help or hinder their performance. In both cases, larger factions tended to win, but the larger factions were more likely to favor the correct alternative for valid syllogisms as compared to invalid syllogism. In addition, two-person minorities favoring the incorrect alternative were very unpersuasive for the valid syllogism, but two-person minorities favoring the incorrect alternative were quite persuasive for the invalid syllogisms. Thus, both group performance trends—above and below the individual baseline—can generally be explained by basic processes associated with shared preferences and shared representations of the task.

Study 2

Study 2 is also an extension of earlier research looking at group performance on tasks where individuals tend to apply simple but potentially inappropriate heuristics to probability estimation problems (Tversky & Kahneman, 1974). The task domain used involved estimating conjoint probabilities, using the classic “Linda the feminist bank teller” problem (Tversky & Kahneman, 1983). Kahneman and Tversky had participants read a short description of Linda, which described her as outgoing, athletic, extroverted, and liberal. Then they asked participants to estimate the likelihood that Linda would fall into various categories. One of the questions asked participants to estimate the likelihood that Linda was a bank teller, which tended to produce low estimates because it was not consistent with the stereotype created by the description. Another question asked for an estimate for Linda being a feminist, which led to relatively high judgments since it did fit the stereotype. A third question asked participants to estimate the likelihood that Linda was a “feminist bank teller”. Since feminist bank teller is, by definition, a subset of bank tellers, probability theory argues that estimates for feminist bank teller should be less than or equal to estimates for bank teller. However, most people judged feminist bank teller as more likely than simply bank teller because it was a better “representation” of the description of Linda provided. Yates and Carlson (1986) found that people generally overestimate conjunction probabilities when one of the component parts is considered fairly likely. When both components are unlikely, the likelihood of making a conjunction error (overestimating the likelihood of the conjunction) is drastically reduced. However, when both components are quite likely, the error rate is increased.

Our initial work comparing individuals and groups using problems modeled after the Linda problem showed that groups were more likely than individuals to make conjunction errors for those types of conjunctions with high individual error rates, but make fewer errors than individuals for conjunctions with low error rates (Tindale

et al., 1996). More recent research has shown that individuals are less likely to overestimate conjunctions when they are familiar with the actual frequencies associated with the components of the conjunction (Gigerenzer & Hofferage, 1996). Thus, it is possible that groups might be less error prone when making judgments about categories that they have experience with or are knowledgeable about. We also had the ability to videotape the group discussions in the current study which we were not able to do in our previous research.

Method

Participants Participants were 470 undergraduate introductory psychology students drawn from two Midwestern universities. Students participated either as individuals (122) or as members of four-person groups. All participants were given course credit for their participation

Materials and Procedures Upon arrival at the experiment, participants were randomly assigned to participate as individuals or members of 4-person groups. All participants then were then asked to read a short paragraph about a fictitious person and to make a series of likelihood judgments concerning that person as individuals. Likelihood judgments were made on 100-point scales where 0 = 0% likely and 100 = 100% likely. The paragraph and estimates were modeled after the “Linda” problem used by Tversky and Kahneman (1983). Half of the participants (those in the unfamiliar condition) received a paragraph describing a generic male person with no particular relationship or relevance to the participants, much like Linda in the feminist bank teller problem. The description of this person implied introversion, intelligence, and high business and mathematical abilities. The likelihood judgments about the person used these traits to define the likely and unlikely categories (e.g., social director at his condominium—unlikely; good chess player—likely). The other half of the participants received a paragraph describing a male student at the participants’ university (familiar condition).

This person was described as a typical student at the university and the likelihood judgments were based on categories that most students would know to be common or rare (e.g., from the Chicago area—likely; a math major—unlikely). Participants made nine likelihood judgments, of which three were used as the main dependent variables. Participants made likelihood judgments for 5 single categories, and also made estimates for three conjunction categories; two likely categories, one likely and one unlikely category, and two unlikely categories. Each of the component parts of the conjunctions were included in the five single category ratings. Thus we had participant ratings for each conjunction and their ratings for each of the component parts of the conjunction.

Following the initial judgments, individual participants worked on an unrelated task, while group participants were asked to make the nine likelihood judgments again but this time as a group. The groups were told that they could make their judgments in any way that they liked but the final group response should reflect the group's collective opinion. After the groups finished making their group judgments and the individuals finished the unrelated task, participants again made individual likelihood judgments for the nine questions. Following this last set of judgments, the participants were debriefed, thanked for their participation, and excused from the experiment.

Results

Individual and group likelihood judgments for the three conjunctions were scored as correct or incorrect by comparing the likelihood judgment for the conjunction to the likelihood judgments for the two components of the conjunction. The judgment was scored "correct" if it was less than or equal to the lower of the two component judgments. Otherwise, it was scored an error. The final individual judgments and group judgments were subjected to a 2 (individual vs. group) by 2 (familiar vs. unfamiliar case) x 3 (type of conjunction: likely–likely (LL), likely–unlikely (LU), and

unlikely–unlikely (UU)) repeated measures analysis of variance. The only effects that reached significance were the main effect for conjunction type, $F(2, 410) = 49.12, p < .001, \text{partial } \eta^2 = .193$; the conjunction type by individual-group interaction, $F(2, 410) = 17.96, p < .001, \text{partial } \eta^2 = .081$; the conjunction type by familiarity interaction, $F(2, 410) = 6.17, p < .01, \text{partial } \eta^2 = .029$; and the conjunction type by individual-group by familiarity interaction, $F(2, 410) = 3.29, p < .05, \text{partial } \eta^2 = .016$. In order to further explore the three-way interaction, we ran three 2 (individual-group) x 2 (familiarity) analyses of variance, one for each conjunction type. It appears that the three-way interaction stems mainly from the fact that there is a nearly significant individual-group by familiarity interaction for the LU conjunction, $F(1, 410) = 2.98, p < .09, \text{partial } \eta^2 = .008$, but not for the UU, $F(1, 410) = .084, p = .77, \text{partial } \eta^2 < .001$, or the LL, $F(1, 410) = .51, p = .48, \text{partial } \eta^2 = .001$.

As shown in Figure 4, groups (88% correct) generally performed better than individuals (72% correct) for the UU conjunctions (panel a), $F(1, 410) = 12.63, p < .001, \text{partial } \eta^2 = .034$, but groups (35% correct) performed worse than individuals (51% correct) for the LL conjunctions (panel b), $F(1, 410) = 8.46, p < .01, \text{partial } \eta^2 = .023$. However, for the LU conjunctions (panel c), groups performed better than individuals only in the familiar condition (groups—80%; individual—59%), $F(1, 410) = 5.06, p < .05, \text{partial } \eta^2 = .047$, and performed slightly worse (though not significantly so) than individuals in the unfamiliar condition (groups—53%; individuals—59%), $F(1, 410) = .31, p = .58, \text{partial } \eta^2 = .003$. Thus, familiarity only seemed to impact group judgments when the unlikely component was combined with a more likely component. Replicating past research, when both components were likely, groups exacerbated the error tendencies found at the individual level, but, when both components were unlikely, groups made fewer errors than individuals.

We once again looked at the group decision processes in two ways. First, as shown in Tables 4, 5, and 6, we calculated the observed social decision schemes for groups in both the familiar and

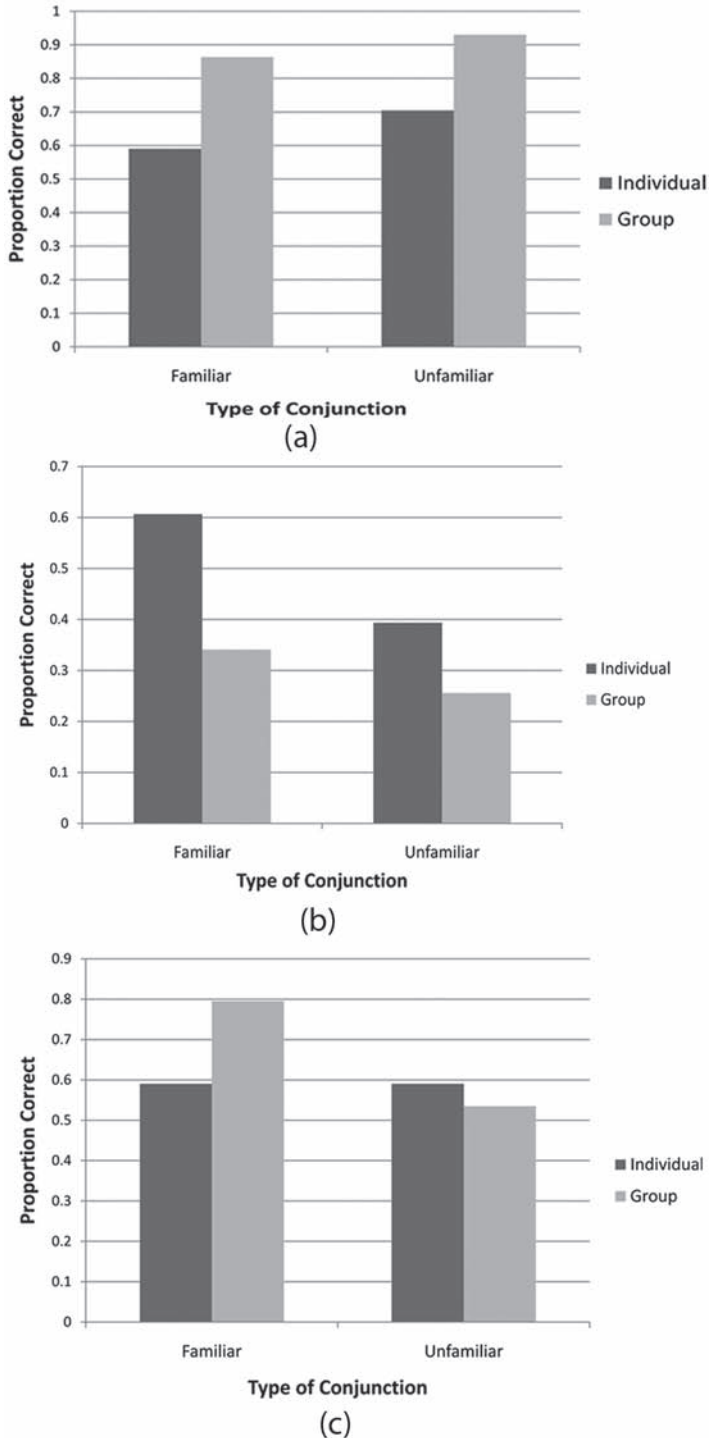


Figure 4. Individual and group performance across familiarity conditions for (a) Unlikely–Unlikely, (b) Likely–Likely, and (c) Likely–Unlikely conjunctions.

Table 4. Observed Social Decision Scheme matrices for both familiar and unfamiliar versions of the Unlikely–Unlikely conjunctions

Individual distributions	Group distributions					
	Familiar			Unfamiliar		
	N	C	I	N	C	I
C–I						
4–0	12	0.92	0.08	15	1.00	0.00
3–1	13	0.92	0.08	15	1.00	0.00
2–2	13	0.85	0.15	10	0.70	0.30
1–3	4	0.75	0.25	2	1.00	0.00
0–4	1	0.00	1.00	0	–	–

C = Correct Response; I = Incorrect Response; N = Number of groups.

Table 5. Observed Social Decision Scheme matrices for both familiar and unfamiliar versions of the Likely–Likely conjunctions

Individual distributions	Group distributions					
	Familiar			Unfamiliar		
	N	C	I	N	C	I
C–I						
4–0	9	0.44	0.568	0	–	–
3–1	9	0.33	0.67	6	0.33	0.67
2–2	15	0.33	0.67	5	0.40	0.60
1–3	9	0.22	0.78	17	0.24	0.76
0–4	2	0.00	1.00	15	0.20	0.80

C = Correct Response; I = Incorrect Response; N = Number of groups.

Table 6. Observed Social Decision Scheme Matrices for both familiar and unfamiliar versions of the Likely–Unlikely conjunctions

Individual distributions	Group distributions					
	Familiar			Unfamiliar		
	N	C	I	N	C	I
C–I						
4–0	15	0.73	0.27	2	0.50	0.50
3–1	12	1.00	0.00	11	0.72	0.28
2–2	12	0.58	0.42	17	0.47	0.53
1–3	3	1.00	0.00	9	0.33	0.67
0–4	1	1.00	0.00	4	0.25	0.75

C = Correct Response; I = Incorrect Response; N = Number of groups.

unfamiliar conditions for each type of conjunction (UU, LL, and LU respectively).³ For the UU conjunctions (both familiar and unfamiliar, see Table 4), there were many more correct than incorrect majorities and correct minorities were often able to win over incorrect minorities. For the LL conjunctions in both familiarity conditions (see Table 5), the patterns were reversed. Across all observed member preference distributions, groups were more likely than not to make an error, and many more groups had majorities that had made errors as individuals. However, for the LU conjunctions (see Table 6), the patterns were somewhat different for the familiar and unfamiliar conditions. For the unfamiliar condition, there were approximately the same number of majorities favoring the correct and incorrect positions, and there was a very slight tendency for incorrect minorities to be more influential than correct minorities. However, for the familiar conjunctions, there were many more groups with correct majorities and correct minorities tended to win. Thus, for the familiar LU conjunctions, both the shared preferences and the shared knowledge of the category frequencies tended to override the general tendency toward the conjunction error.

We also videotaped the group discussions and two independent coders coded them in terms of how groups seemed to reach consensus ($\alpha = .69$).⁴ Although five different categories emerged from the coding, two of the categories encompassed 86% of the groups that could be coded; group went with a single member’s response (60%), and group chose a compromise or rough combination of two members’ responses (26%).⁵ Table 7 shows the percentage of correct responses for both familiar and unfamiliar conditions for each conjunction type for groups categorized into these two consensus process categories. As can be seen in the table, these consensus processes led to fairly good performance for the UU conjunctions regardless of familiarity. These same processes led to fairly poor performance for both familiar and unfamiliar LL conjunctions. However, for the LU conjunctions, the processes led to fairly good performance in the familiar condition, but not for the unfamiliar

Table 7. Proportion of correct group responses for groups using one of two main judgment strategies by type of conjunction and familiarity

Type of conjunction	Decision maker	Familiar	Unfamiliar
Unlikely–Unlikely	Individual	.61	.70
	Group	.86	.93
Likely–Likely	Individual	.41	.39
	Group	.34	.26
Likely–Unlikely	Individual	.59	.59
	Group	.80	.55

condition, although the differences between the familiarity conditions were only marginally significant, $\chi^2(1, N = 29) = 3.25, p < .10, \Phi = .33$. Thus, seemingly the same processes led to considerably different performance levels across the three types of conjunctions, and for the LU conjunctions, led to performance differences as a function of familiarity.

General discussion

As previously noted, it is often assumed by both the popular press and many groups researchers that good vs. poor group performance varies as a function of good vs. poor group processes (e.g., Janis, 1982). Although there is evidence that certain group process characteristics can lead to better or worse performance (Brodbeck et al, 2007; De Dreu, Nijstad, & von Kippenberg, 2008) our general argument is that most groups probably do not display either particularly good or poor process, but rather typically function under fairly basic social processes, one of which we refer to as social sharedness (Tindale & Kameda, 2000). Social sharedness tends to serve groups well in many common decision domains, as exemplified by the finding that groups often outperform individuals (Davis, 1969; Kerr & Tindale, 2011; Larson, 2010). However, in situations where shared knowledge is biased or inappropriate for a particular task, groups will tend to share preferences that are incorrect or suboptimal, yet in the group discussion, such preferences will seem plausible even to members who did not favor them at the beginning of the discussion. In these situations, groups doing much of what they normally do will not produce the

outcomes (superior performance) we typically expect from groups.

The results of Study 1 showed that groups can both perform well and poorly on syllogistic reasoning tasks depending on whether their normative assumptions interfere with the formal logic underlying the problem. For those syllogisms where the conversational norm “stating the qualifier some implies that the categories only partially overlap” did not interfere with judging the syllogisms validity, groups did perform better than individuals. However, for those syllogisms where the norm did interfere, the groups performed worse than individuals. Study 2 showed similar performance patterns for judging the likelihood of conjunctive events. In those situations where the members’ basic tendencies led to lower judgments (e.g., when both components were unlikely) groups performed better than individuals. However, when those basic tendencies led to high judgments (e.g., when both component likely), groups performed more poorly than individuals. Extending our previous research in this area, we also found that using categories with which the group members were familiar allowed groups, to a greater degree than individuals, to avoid conjunction errors when conjoining a likely and an unlikely category. It could be that the direct knowledge of the size of an unlikely category makes it easier to construe the conjunction more appropriately and to make lower judgments seem more plausible. This greater plausibility may also affect how confident a particular member is in terms of his/her initial estimate, which could increase the speed and intensity of how the estimate is presented to the group.

Even when we looked at group processes at a more fine grained level, we found that both good and poor performance could be produced by similar processes. For the syllogism problems, discussing the meaning or implications of premises and conclusions that contained the qualifier “some” always led to the resolution that some implies some are not. This extra scrutiny may have implied that the groups were attempting to use logic to solve the problems. When the resolution did not interfere with the logic of the problem, the discussion appeared to improve performance. However, when the resolution did interfere, it led to poorer performance. For the conjunction task, two dominant consensus processes emerged: choose one members preference or merge preferences from two members. These processes were equally likely across all conditions, yet led to good performance for the UU conjunctions and poor performance for the LL conjunctions. In essence, many of our groups in both studies were doing the same things, but those things were beneficial in some cases and not beneficial in others.

There are two possible conclusions one could take away from the research described here that would be incorrect.⁶ First, just because some of our groups performed rather poorly should not be used to argue that using groups to make important decisions is bad. The vast majority of research on group decision making and problem solving argues exactly the opposite: groups very often perform better, and rarely perform worse, than the level of performance one would expect by a single individual (Larson, 2010). All our results show is that groups can make mistakes even when they are acting in ways that typically lead to good performance. In other words, groups do not have to fall prey to poor decision practices in order to perform poorly. Second, it would also be erroneous to assume that groups cannot be taught to perform in more optimal ways. Groups, on their own, will often perform well, but group training to insure good communication, trust, and an accurate shared model of the task and the relevance of the various member roles can produce substantial performance increments (Cannon-Bowers, Salas, & Converse, 1993;

Weiner, Kanki, & Helmreich, 1993). And training groups to watch out for task specific errors can be used to help groups adapt to environments where their typical responses are not effective (Paulus, Nakui, Putman, & Brown, 2006; Semmer, Tschan, Hunziker, & Marsch, 2011). Thus, although groups will typically do well, they can do better. We hope the studies presented here help to focus attention away from over-generalizations from past work (e.g., Janis, 1982) and to help to better isolate what factors truly are important for group performance.

Notes

The research reported here was funded by the following grants from the National Science Foundation (SBR #9730822, SES #0136332, BCS #0621632, & BCS #0820344). Portions of this article were drawn from the Midwestern Psychological Association’s Presidential Address (2008) by the first author. The authors represent four generations of James Davis’ professional progeny as each later author began their careers as a student of the preceding author and the first author was a student of Jim’s

- 1 An analysis using the sum of the number of correct responses (any rating of 5 or greater for valid syllogisms and 4 or less for invalid syllogisms) produced the identical interaction and simple effects results.
- 2 Due to technical issues, particularly poor sound quality, only 42 of the groups produced discussions that could be reliably coded.
- 3 The numbers of groups per row within each matrix in each condition were relatively small, so we did not attempt statistical comparisons between conditions.
- 4 Once again, technical difficulties limited to 41 the number of group discussions that could be reliably coded. Since each group estimated likelihoods for three conjunctions, we had 122 cases (one group could not be coded for the LU conjunction) .
- 5 The other three categories were: Averaged the two component scores of the conjunction (6%), correctly insured that the conjunction estimate was below both components (7%), and multiplied the two component estimates together (< 1%). Because of their low numbers and because each of these strategies, by definition, leads either to a

correct response or an error, we did not include them in the reported analyses.

- 6 We bring these up here because they are conclusions that have been inferred by audience members when we have presented these results at conferences.

References

- Baron, R. S. (2005). So right it's wrong: Groupthink and the ubiquitous nature of self censorship. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 37, pp. 219–253). San Diego, CA: Elsevier.
- Brewer, M. B., & Capoeael, L. R. (2006). An evolutionary perspective on social identity: Revisiting groups. In M. Schaller, J. A. Simpson, & D. T. Kenrick (Eds.) *Evolution and social psychology* (pp. 143–162). New York, NY: Psychology Press.
- Brodbeck, F. C., Kerschreiter, R., Mojzisch, A., Frey, D., & Schulz-Hardt, S. (2007). Group decision making under conditions of distributed knowledge: The information asymmetries model. *Academy of Management Journal*, *32*, 459–479. <http://search.proquest.com/docview/210989238?accountid=12163>
- Cannon-Bowers, J. A., Salas, E., & Converse, S. (1993). Shared mental models in expert shared decision making. In N. Castellan, Jr., (Ed.) *Individual and group decision making: Current issues* (pp. 221–246). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Davis, J. H. (1969). *Group performance*. New York, NY: Addison-Wesley.
- Davis, J. H. (1973). Group decisions and social interactions: A theory of social decision schemes. *Psychological Review*, *80*, 97–125. doi: apa.org/psycinfo/1973-20900-001
- Davis, J. H., Kerr, N. L., Atkin, R. S., Holt, R., & Meek, D. (1975). The decision processes of 6- and 12-person juries assigned unanimous or 2/3 majority rules. *Journal of Personality and Social Psychology*, *32*, 1–14. doi: [10.1037/h0076849](https://doi.org/10.1037/h0076849)
- De Dreu, C.K.W., Nijstad, B.A., & Van Knippenberg, D. (2008). Motivated information processing in group judgment and decision making. *Personality and Social Psychology Review*, *12*, 22–49. doi: [10.1177/1088868307304092](https://doi.org/10.1177/1088868307304092)
- Evans, J. St.B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning: The psychology of deduction*. East Sussex: Lawrence Erlbaum.
- Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychology Review*, *102*, 684–704. doi: [10.1037/0033-295X.102.4.684](https://doi.org/10.1037/0033-295X.102.4.684)
- Grice, P. (1975). Logic and conversation. In P. Cole, & J. L. Morgan (Eds.) *Studies in syntax, Vol. 3: Syntax*. New York, NY: Academic Press.
- Hackman, J. R. (1998). Why groups don't work. In R. S. Tindale, L. Heath, J. Edwards, E. J. Posavac, F. B. Bryant, Y. Suarez-Balcazar, E. Henderson-King, & J. Myers (Eds.) *Social psychological applications to social issues: Theory and research on small groups* (vol. 4, pp. 245–268). New York, NY: Plenum Press.
- Hill, G. W. (1982). Group vs. individual performance: Are N + 1 heads better than one? *Psychological Bulletin*, *91*, 517–539. doi: [0033-2909/82/9103-0517S00.75](https://doi.org/10.1037/0033-2909.82.9103-0517S00.75)
- Hinsz, V. B., Tindale, R. S., & Nagao, D. H. (2008). Accentuation of information processes and biases in group judgments integrating base-rate and case-specific information. *Journal of Experimental Social Psychology*, *44*, 116–126. <http://dx.doi.org/10.1016/j.jesp.2007.02.013>
- Hinsz, V. B., Tindale, R. S., & Vollrath, D. A. (1997). The emerging conception of groups as information processors. *Psychological Bulletin*, *121*, 43–64. doi: [10.1037/0033-2909.121.1.43](https://doi.org/10.1037/0033-2909.121.1.43)
- Janis, I. (1982). *Groupthink* (2nd ed.). Boston, MA: Houghton-Mifflin.
- Kameda, T., & Tindale, R. S. (2006). Groups as adaptive devices: Human docility and group aggregation mechanisms in evolutionary context. In M. Schaller, J. A. Simpson, & D. T. Kenrick (Eds.) *Evolution and social psychology* (pp. 317–342). New York, NY: Psychology Press.
- Kameda, T., Tindale, R. S., & Davis, J. H. (2003). Cognitions, preferences, and social sharedness: Past, present and future directions in group decision making. In S. L. Schneider, & J. Shanteau (Eds.) *Emerging perspectives on judgment and decision research* (pp. 458–485). New York, NY: Cambridge University Press.
- Kameda, T., Wisdom, T., Toyodawa, W., & Inukai, K. (2012). Is consensus seeking unique to humans? A selective review of animal group decision-making and its implications for (Human) social psychology. *Group Processes and Intergroup Relations*, *15*(5), 673–689.
- Kerr, N. L., & Tindale, R. S. (2004). Small group decision making and performance. *Annual Review of Psychology*, *55*, 623–656. doi: [10.1146/annurev.psych.55.090902.142009](https://doi.org/10.1146/annurev.psych.55.090902.142009)
- Kerr, N. L., & Tindale, R. S. (2011). Group-based forecasting: A social psychological analysis. *International*

- Journal of Forecasting*, 27, 14–40. doi:10.1016/j.ijforecast.2010.02.001
- Larson, J. R. (2010). *In search of synergy in small group performance*. New York, NY: Psychology Press.
- Laughlin, P. R. (1980). Social combination processes of cooperative, problem-solving groups on verbal intellectual tasks. In M. Fishbein (Ed.), *Progress in social psychology* (Vol. 1, pp. 127–155). Hillsdale, NJ: Lawrence Erlbaum.
- Laughlin, P. R. (1999). Collective induction: Twelve postulates. *Organizational Behavior and Human Decision Processes*, 80, 50–69. <http://dx.doi.org/10.1006/obhd.1999.2854>
- Laughlin, P. R., & Ellis, A. L. (1986). Demonstrability and social combination processes on mathematical intellectual tasks. *Journal of Experimental Social Psychology*, 22, 177–189. [http://dx.doi.org/10.1016/0022-1031\(86\)90022-3](http://dx.doi.org/10.1016/0022-1031(86)90022-3)
- Lorge, I., & Solomon, H. (1955). Two models of group behavior in the solution of eureka-type problems. *Psychometrika*, 20, 139–148. doi: 10.1007/BF02288986
- Mullen, B., & Copper, C. (1994). The relation between group cohesiveness and performance: An integration. *Psychological Bulletin*, 115(2), 210–227. doi: 10.1037/0033-2909.115.2.210
- Nijstad, B. A. (2009). *Group performance*. New York, NY: Psychology Press.
- Paulus, P. B., Nakui, T., Putman, V. L., & Brown, V. R. (2006). Effects of task instructions and brief breaks on brainstorming. *Group Dynamics: Theory, Research, and Practice*, 10, 206–219. doi: 10.1037/1089-2699.10.3.206
- Peterson, R. S. (1997). A directive leadership style in group decision making is both virtue and vice: Evidence from elite and experimental groups. *Journal of Personality and Social Psychology*, 72, 1107–1121. doi: 10.1037/0022-3514.72.5.1107
- Postmes, T., Spears, R. and Cihangir, S. (2001). Quality of decision making and group norms. *Journal of Personality and Social Psychology*, 80, 918–930. doi: 10.1037/0022-3514.80.6.918
- Semmer, N.K., Tschan, F., Hunziker, S., & Marsch, S. U. (2011). Leadership and minimally invasive training enhance performance in medical emergency driven teams: Simulator studies. In V. G. Duffy, (Ed.). *Advances in human factors and ergonomics in healthcare* (pp. 180–190). Boca Raton, CA: Taylor & Francis.
- Smith, C. M., Tindale, R. S., Kameda, T., Dykema-Engblade, A. Niven, T. S., Krebel, A., Munier, C., & Nakanishi, D. (2000). *Shared representations and group performance on syllogistic reasoning problems: A cross cultural investigation*. Paper presented at the Nags Head Conference on Groups, Networks, and Organizations, Highland Beach, FL.
- Smith, C. M., Tindale, R. S., & Steiner, L. (1998). Investment decisions by individual and groups in “Sunk Cost” situations: The potential impact of shared representations. *Group Processes and Intergroup Relations*, 1, 175–189. doi: 10.1177/1368430298012005
- Tindale, R. S. (1993). Decision errors made by individuals and groups. In N. Castellan, Jr., (Ed.) *Individual and group decision making: Current issues* (pp. 109–124). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Tindale, R. S., Davis, J. H., Vollrath, D. A., Nagao, D. H., & Hinsz, V. B. (1990). Asymmetrical social influence in freely interacting groups: A test of three models. *Journal of Personality and Social Psychology*, 58, 438–449. doi: 10.1037/0022-3514.58.3.438
- Tindale, R. S., & Kameda, T. (2000). Social sharedness as a unifying theme for information processing in groups. *Group Processes and Intergroup Relations*, 3, 123–140. doi: 10.1177/1368430200003002002
- Tindale, R. S., Smith, C. M., Thomas, L. S., Filkins, J., & Sheffey, S. (1996). Shared representations and asymmetric social influence processes in small groups. In E. Witte, & J. Davis (Eds.) *Understanding group behavior: Consensual action by small groups* (Vol. 1, pp. 81–103). Mahwah, NJ: Lawrence Erlbaum Associates.
- Tindale, R. S., Talbot, M., & Martinez, R. (in press). Group decision making. In J. Levine (Ed.) *Frontiers in social psychology: Group processes*. New York, NY: Psychology Press.
- Tversky, A., & Kahneman, D. (1974). Judgments under uncertainty: Heuristics and biases. *Science*, 185, 1124–1131. doi: 10.1126/science.185.4157.1124
- Tversky, A., & Kahneman, D. (1980). Causal schemas in judgments under uncertainty. In M. Fishbein (Ed.) *Progress in social psychology* (Vol. 1, pp. 49–72). Hillsdale, NJ: Lawrence Erlbaum.
- Tversky, A., & Kahneman, D. (1983). Extensional vs. intuitive reasoning: The conjunction fallacy in probability judgments. *Psychological Review*, 90, 283–315. doi: 10.1037/0033-295X.90.4.293
- Wegner, D.M. (1987). Transactive memory: A contemporary analysis of the group mind. In B. Mullen & G.R. Goethals (Eds.), *Theories of group*

- behavior* (pp. 185–208). New York, NY: Springer-Verlag.
- Weiner, E. L., Kanki, B., & Helmreich, R. L. (1993). *Cockpit resource management*. San Diego, CA: Academic Press
- Yates, J. F., & Carlson, B. W. (1986). Conjunction errors: Evidence for multiple judgment procedures including “signed summation”. *Organizational Behavior and Human Decision Processes*, 37, 230–253. [http://dx.doi.org/10.1016/0749-5978\(86\)90053-1](http://dx.doi.org/10.1016/0749-5978(86)90053-1)