Social Loafing: A Meta-Analytic Review and Theoretical Integration

Steven J. Karau and Kipling D. Williams

Social loafing is the tendency for individuals to expend less effort when working collectively than when working individually. A meta-analysis of 78 studies demonstrates that social loafing is robust and generalizes across tasks and S populations. A large number of variables were found to moderate social loafing. Evaluation potential, expectations of co-worker performance, task meaningfulness, and culture had especially strong influence. These findings are interpreted in the light of a Collective Effort Model that integrates elements of expectancy-value, social identity, and self-validation theories.

Many of life's most important tasks can only be accomplished in groups, and many group tasks are collective tasks that require the pooling of individual members' inputs. Government task forces, sports teams, organizational committees, symphony orchestras, juries, and quality control teams provide but a few examples of groups that combine individual efforts to form a single product. Because collective work settings are so pervasive and indispensable, it is important to determine which factors motivate and demotivate individuals within these collective contexts. Intuition might lead to the conclusion that working with others should inspire individuals to maximize their potential and work especially hard. Research on social loafing, however, has revealed that individuals frequently exert less effort on collective tasks than on individual tasks.

Formally, *social loafing* is the reduction in motivation and effort when individuals work collectively compared with when they work individually or coactively. When working collectively, individuals work in the real or imagined presence of others with whom they combine their inputs to form a single group product. When working coactively, individuals work in the real or imagined presence of others, but their inputs are not combined with the inputs of others. Determining the conditions under which individuals do or do not engage in social loafing is a problem of both theoretical and practical importance. At a practical level, the identification of moderating variables may suggest means for devising interventions by which social loafing may be reduced or overcome in everyday groups and organizations. Latané, Williams, & Harkins (1979) even suggested that social loafing is a type of social disease, having "negative consequences for individuals, social institutions, and societies" (p. 831). Perhaps as a result of this characterization, social loafing research has been focused on identifying conditions under which the effect can be reduced or eliminated. No studies have been designed to determine what factors increase social loafing. This emphasis on eliminating the effect is understandable, given the potential applicability of social loafing research to real-world contexts. However, this emphasis on studying conditions in which the effect is not likely to occur may also result in an underestimation of the magnitude of social loafing across a wider range of situations.

At a theoretical level, specifying which variables moderate social loafing is central to developing a fuller understanding of the dynamics underlying the performance and motivation of both individuals and groups. The social loafing literature offers a host of findings relevant to theorists interested in evaluation processes, the self, and group dynamics. Indeed, in a recent review of social motivation, Geen (1991) regarded social loafing as one of only three dominant phenomena addressing this issue in the 1980s. Despite the theoretical importance of this topic, surprisingly little attention has been devoted to systematically reviewing or integrating the social loafing research. Although a number of researchers have discussed social loafing or have presented theories of particular causes of social loafing (e.g., Harkins, 1987; Harkins & Szymanski, 1987; Jackson & Williams, 1985; Latané, 1981; Mullen, 1983; Paulus, 1983; Shepperd, 1993; Stroebe & Frey, 1982), they have drawn their conclusions from only small portions of the available empirical studies, which were selected by unspecified criteria. Presently, the magnitude and consistency of social loafing across studies has never been estimated, and the extent to which particular variables moderate social loafing has remained unclear. Given the large empirical literature that is now available, a thorough integration and analysis of this research is long overdue.

Parts of this article were presented at the 1992 Midwestern Psychological Association convention in Chicago.

We thank Martin Bourgeois, Donal Carlston, Alice Eagly, Rebecca Henry, Janice Kelly, Norbert Kerr, Brian Mullen, Kristin Sommer, and several anonymous reviewers for their valuable comments on drafts of this article. We also thank Jeffrey Jackson and William Gabrenya for providing unpublished data.

Correspondence should be addressed to Steven J. Karau, Department of Psychology, Clemson University, Clemson, South Carolina 29634-1511 or to Kipling D. Williams, Department of Psychology, University of Toledo, Toledo, Ohio 43606-3390. Electronic mail may be sent to skarau@clemson.edu or kipling@uoft02.utoledo.edu.

Journal of Personality and Social Psychology, 1993, Vol. 65, No. 4, 681-706
Copyright 1993 by the American Psychological Association, Inc. 0022-3514/93/$3.00

The current review has three major purposes. First, we present a model that integrates expectancy theories with theories of group-level social comparison and social identity to account for the findings of studies examining individual effort in collective settings. This model, the Collective Effort Model (CEM), is then used to generate predictions for the meta-analysis. Second, we systematically review the available empirical findings. Meta-analytic methods (Cooper, 1989; Glass, McGaw, & Smith, 1981; Hedges & Olkin, 1985; Rosenthal, 1991) are used to estimate the overall magnitude of social loafing and to compare it with the magnitude of the effects of other variables on social behavior. This quantitative method also allows for an examination of the consistency of social loafing across studies and accounts for inconsistencies by identifying moderating variables that affect the tendency for individuals to engage in social loafing. The empirical status of such moderating variables is critical to a sophisticated understanding of the dynamics of motivation losses in groups. Third, we use the results of the meta-analysis to identify gaps and ambiguities still existing in the social loafing literature and discuss the potential of the CEM for generating novel predictions to guide future research toward answering these questions.

Brief History of Social Loafing Research and Paradigms

The first experiment to suggest a possible decrement in individual motivation as a result of working in a group was conducted over a hundred years ago by Ringelmann (cited in Kravitz & Martin, 1986). Male volunteers were asked to pull on a rope, tug-of-war fashion, as hard as they could in groups of varying sizes. The rope was connected to a strain gauge that measured the group's total effort. The results showed that as group size increased, group performance was increasingly lower than would be expected from the simple addition of individual performances. However, Steiner (1972) proposed two possible causes for this performance decrement: (a) reduced individual motivation or (b) coordination loss. Steiner favored the latter cause as most parsimonious, concluding that individuals may fail to synchronize their efforts in a maximally efficient manner (e.g., pulling while others are pausing), thus evidencing less productivity, but not necessarily less effort.

To separate effort reduction from coordination loss, Ingham, Levinger, Graves, and Peckham (1974) had subjects perform a rope-pulling task both in actual groups and in pseudogroups in which blindfolded male students believed they were pulling with other group members but were actually pulling alone. Data for the pseudogroup trials showed that performance still decreased as perceived group size increased, suggesting that individuals exerted less effort when working in groups than when working individually. However, group size covaried inversely with audience size in Ingham et al.'s studies, leaving open the possibility that effort was facilitated by larger audiences in the individual trials.

Latané et al. (1979) conceptually replicated these findings while holding audience size constant. They had subjects shout and clap as loudly as they could, both individually and with others. Blindfolded and wearing headphones that masked the noise, college men shouted both in actual groups and in pseudogroups in which they shouted alone but believed they were shouting with others. Latané et al. demonstrated that a substantial portion of the decreased performance of groups was attributable to reduced individual effort, distinct from coordination loss, and that audience size did not account for these results. They also coined the term *social loafing* for the demotivating effects of working in groups.

Since 1974, nearly 80 studies on social loafing have been conducted in which individuals' coactive efforts were compared with individuals' collective efforts. These studies have used a wide variety of tasks, including physical tasks (e.g., shouting, rope-pulling, and swimming), cognitive tasks (e.g., generating ideas), evaluative tasks (e.g., quality ratings of poems, editorials, and clinical therapists), and perceptual tasks (e.g., maze performance and vigilance tasks on a computer screen). Both laboratory experiments and field studies have been conducted using a range of subject populations varying in age, gender, and culture.

Theoretical Accounts for Social Loafing

Although several researchers have offered theories of social loafing, these viewpoints have typically been restricted to explaining only one of several possible causal mechanisms and generally do not attempt to include the wide range of variables that moderate social loafing. In contrast, social loafing research has adopted an orientation toward isolating conditions under which the effect can be reduced or eliminated. As a result, despite the tendency of researchers to manipulate a wide array of moderating variables across individual studies, there has not been a concerted attempt to integrate all of these moderators into a single theoretical model. Indeed, each moderating variable is frequently associated with a separate theory. Nevertheless, the theories that have been advanced do provide some insight into the nature and possible causes of social loafing. Therefore, before presenting the CEM, we first present an overview of viewpoints offered by prior researchers and discuss the need for an integrative model.

Social Impact Theory

According to Latané's (1981) social impact theory, people can be viewed as either sources or targets of social impact. The amount of social impact experienced in a situation is thought to be a function of the strength, immediacy, and number of sources and targets present. In the collective condition of the typical social loafing experiment, the experimenter serves as a single source of social impact, whereas the group members usually serve as multiple targets of social impact. Social impact theory suggests that the experimenter's request to try as hard as possible on the task should be divided across targets, resulting in reduced effort as group size increases. The division of impact is predicted to follow an inverse power function, with a negative exponent having an absolute value less than 1, thereby resulting in marginally decreasing impact as group size increases. In the coactive condition, however, in which inputs are not combined, each individual feels the full impact of the experimenter's request and works hard. The strength of social impact theory appears to be its ability to specify group size effects. The strength and immediacy factors, however, have been neglected

by researchers, and the theory itself has been criticized for not addressing the underlying psychological processes that it describes (e.g., Mullen, 1985).

Arousal Reduction

Jackson and Williams (1985) proposed a drive explanation to accompany a social impact theory explanation of social loafing. They argued that the presence of other co-workers is drive reducing because these others serve as cotargets of an outside source of social impact. Citing studies in which people faced with a fearful situation tended to prefer being in the company of others (e.g., Schachter, 1959; Wrightsman, 1960), they reasoned that the presence of others is not necessarily drive inducing (as has been repeatedly demonstrated in social facilitation research). Instead, the presence of others should only be drive inducing when those others serve as sources of impact, but should be drive reducing when they serve as cotargets. Jackson and Williams (1985) found support for this logic in an experiment that combined features of the social loafing and social facilitation paradigms. Subjects completed simple and complex computer mazes either alone, coactively, or collectively. On simple tasks, subjects performed better coactively than collectively. On complex tasks, however, subjects performed better collectively than coactively. Jackson and Williams argued that working collectively led to reduced drive and effort, resulting in decreased performance on simple tasks (where the dominant response is likely to be correct) and increased performance on novel, difficult tasks (where the dominant response is likely to be in error).

Evaluation Potential

Many interpretations of social loafing invoke the concept of evaluation potential (e.g., Harkins, 1987; Harkins & Jackson, 1985; Williams et al., 1981). Some researchers (Harkins, 1987; Harkins & Szymanski, 1987, 1989; Kerr & Bruun, 1983) have even defined social loafing as motivation loss in groups caused by reduced identifiability or evaluation.[1] They argue that social loafing occurs because, in most studies, individuals' inputs can only be evaluated in the coactive condition. In the collective condition, of course, individual inputs are combined into one group product. When working on collective tasks, individuals can "hide in the crowd" (Davis, 1969) and avoid taking the blame for a poor group performance. Collective tasks may also lead individuals to feel "lost in the crowd" (Latané et al., 1979) such that they cannot receive their fair share of the credit for a good group performance. The role of identifiability in social loafing is illustrated in a study by Williams, Harkins, and Latané (1981). Individuals performed a shouting task coactively and collectively. Evaluation potential was manipulated such that individual inputs were either always identifiable, never identifiable, or identifiable only in the coactive condition. Individuals only loafed when identifiability and the coactive–collective variable were confounded. Recent research by Harkins and colleagues (Harkins, 1987; Harkins & Jackson, 1985; Harkins & Szymanski, 1987, 1989) has suggested that making individuals' collective inputs evaluable to anyone (including oneself) may be enough to eliminate social loafing in many situations. This research has also demonstrated that two requirements must be met for evaluation by any source (the experimenter, one's co-workers, or oneself) to be possible: (a) the participant's output must be known or identifiable, and (b) there must be a standard (personal, social, or objective) with which this output can be compared.

Dispensability of Effort

Kerr and his colleagues (e.g., Kerr, 1983; Kerr & Bruun, 1983) have suggested another possible cause of social loafing: Individuals may exert less effort when working collectively because they feel that their inputs are not essential to a high-quality group product (i.e., are dispensable). In their "free rider" paradigm, these researchers have found that individuals tend to reduce their collective efforts when working on threshold tasks that use a disjunctive rule whereby if any one of the group members reaches a certain performance criterion, the group succeeds and further effort is unnecessary. Moreover, this reduction in individual effort occurs even though each group member's contribution is made identifiable to themselves, their partner, and the experimenter. Thus, on some tasks individuals may be unwilling to exert effort if they feel their input will have little impact on the resulting group product.

Matching of Effort

Jackson and Harkins (1985) proposed that people tend to match their co-workers' efforts when working collectively. According to this position, social loafing occurs because individuals expect others to slack off in groups and, therefore, reduce their own efforts to maintain equity. In Jackson and Harkins's (1985) experiment, participants' expectations of how hard their co-worker would work were manipulated such that participants expected their co-worker either to try hard or not to try hard on a shouting task. Social loafing was eliminated and participants matched their co-worker's anticipated effort. However, in this study, the confederate co-worker's statement of her intended effort was confounded with her evaluation of the experiment's worth. In the high-effort condition, she said that she thought the experiment was interesting and that she was going to try hard. In the low-effort condition, however, she said that she thought the experiment was boring, and that she was not going to try hard. Research on job attitudes (see Zalesny & Ford, 1990) suggests that workers' perceptions of and motivations toward their task are highly influenced by co-workers' task assessments. Thus, the confederate's assessment of the worth of the task could have alone accounted for Jackson and Harkins's

[1] Although evaluation is likely to play an important role in social loafing, it seems premature to define social loafing in these terms because (a) defining the phenomenon in terms of its causes prevents, by definition, the discovery of new causes for the same effect, and (b) evidence suggests that there are other causes of social loafing, such as redundancy of contributions (Harkins & Petty, 1982) and dispensability of effort (Kerr & Bruun, 1983). Therefore, the present review adopts a less restrictive definition of social loafing consistent with the original formulation: the tendency to reduce one's effort when working collectively compared with coactively on the same task.

results. Moreover, recent research by Williams and Karau (1991) has demonstrated that individuals may actually increase their collective effort when they expect their co-workers to perform poorly on a meaningful task, an effect referred to as *social compensation.*

Self-Attention

Mullen (1983) proposed that self-attention underlies social loafing. According to this self-attention perspective, working on a collective task leads to a decrease in self-awareness, leading individuals to disregard salient performance standards and to engage in less self-regulation. Thus, collective performance is lower than coactive performance, because participants are more attentive to task demands and performance standards when working alone. Mullen has found evidence consistent with the notion that self-attention may influence aspects of social behavior in various meta-analyses (see Mullen, 1991, for a summary), but there is currently no evidence that reduced self-attention causes social loafing (see Jackson, 1986). Only one study (Stevenson, 1990) has manipulated self-attention across coactive and collective conditions, and the results did not support the self-attention interpretation. Therefore, although the self-attention hypothesis is provocative, it is in need of empirical support.

Conclusions

Prior theories advanced to explain social loafing all appear to have some limitations. The most significant limitation common to all of these viewpoints is that they tend to offer explanations and make predictions about conditions under which social loafing will occur within a limited domain. For example, Harkins (1987) offered an excellent analysis of the impact of evaluation on social loafing, and Kerr (1983) contributed an equally effective analysis of the role that dispensability plays. However, although current theories of social loafing provide insights as to why the effect occurs at all, they do not provide a framework that can clearly specify which particular factors should moderate social loafing under different conditions. When possible in the current review, we quantitatively examine the adequacy of some of these viewpoints for explaining social loafing. More important, we present a unified theory that proposes to envelop most if not all of the findings in the existing social loafing literature and that offers specific predictions for future research.

Integrative Model of Individual Effort on Collective Tasks

The integrated model of individual effort on collective tasks, the CEM, adapts individual-level expectancy-value models of effort to collective contexts to highlight the most likely threats to motivation and uses recent theories of self-evaluation in group contexts to clarify which outcomes are likely to be valued by individuals when working collectively. The resulting model provides a framework for clearly specifying the implications of any key attribute of a given collective setting (e.g., group size and nature of task) for the motivation of individuals in that setting.

We propose that expectancy-value models of effort (e.g., Heckhausen, 1977; Porter & Lawler, 1968; Vroom, 1964) can be extended to collective contexts by specifying how working in a group influences individuals' perceptions of the relationship between their effort and their expected outcomes. We believe that social loafing occurs because there is usually a stronger perceived contingency between individual effort and valued outcomes when working individually. When working collectively, factors other than the individual's effort frequently determine performance, and valued outcomes are often divided among all of the group members. The notion that expectancy-value models can be applied to motivation in group contexts is not entirely new. For example, several researchers have implicitly used expectancy-value logic to cast light on the findings of social dilemma and social loafing studies (e.g., Kerr, 1983; Olson, 1965; Shepperd, 1993; Williams & Karau, 1991). Naylor, Pritchard, and Ilgen (1980) have also suggested that expectancy-value logic can be extended into a model that can account for organization-level motivational phenomena. Similarly, several researchers have discussed the potential implications of self-evaluation for individual motivation (e.g., Goethals & Darley, 1987; Taylor & Brown, 1988; Tesser, 1988). The CEM, however, is unique in specifying additional contingencies between effort and outcomes that are distinctive to collective contexts, and in using group-level social comparison and self-evaluation theories to specify which factors should influence individual motivation in collective contexts. This combination allows for the clear specification of potential threats to collective motivation in a way that is not restricted to mere task-performance outcomes.

The key features of the CEM are shown in Figure 1. The model suggests that individuals will be willing to exert effort on a collective task only to the degree that they expect their efforts to be instrumental in obtaining valued outcomes. Thus, a number of factors must be perceived as existing before individuals will be willing to exert high levels of effort. Individual effort must relate to individual performance, which must in turn have some impact on the group's performance. The group's performance must then lead to a favorable group outcome, which must be related to a favorable individual outcome. If the task performance setting disrupts any of these relationships in a way that is salient to individuals, then they will not be likely to view their efforts as useful and will not work as hard on the task. Similarly, individuals will not work as hard when the available outcomes are not valued, even if these outcomes are directly related to individual effort. Relevant individual outcomes include things such as objective outcomes (such as pay), self-evaluation information, and feelings of purpose or belonging in one's group. The relative value of these outcomes depends on a number of factors, including the meaningfulness and intrinsic value of the task, the task's importance to the individual and group, the degree to which the individual is dispositionally predisposed to view collective outcomes as important, and the degree to which the outcome provides information relevant to the individual's self-evaluation.

Like traditional expectancy-value models of effort, the CEM assumes that individuals behave hedonistically and try to maximize the expected utility of their actions. In Vroom's (1964) original model, individuals' motivational force is dependent on

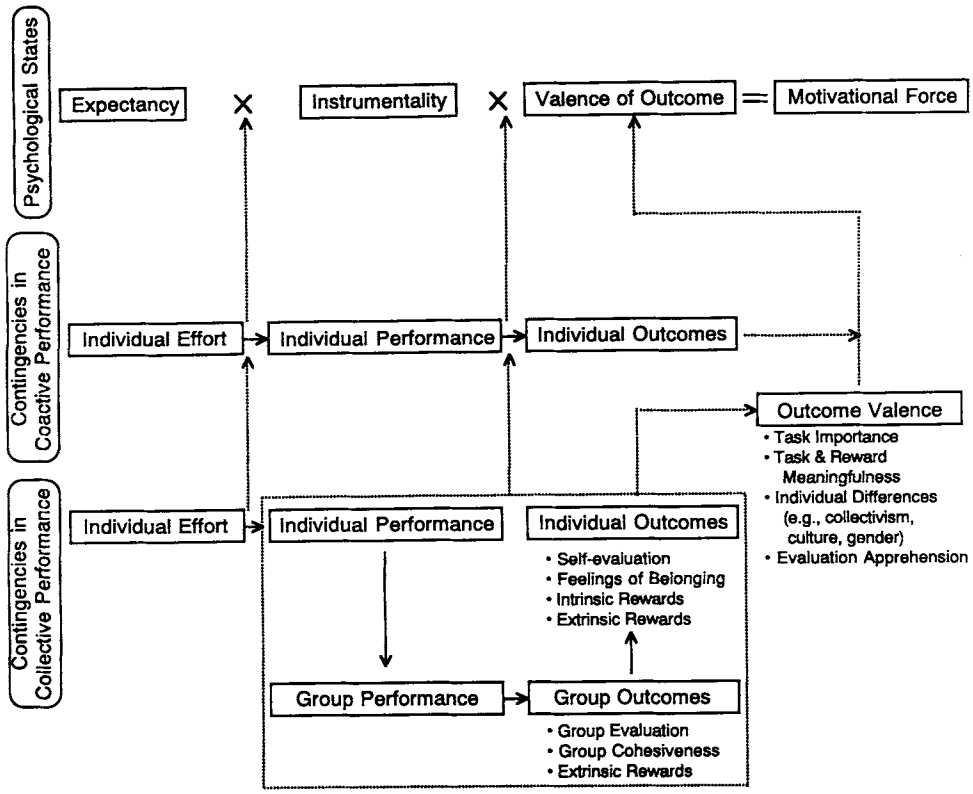

Figure 1. The Collective Effort Model (CEM).

three factors: (a) *expectancy,* or the degree to which high levels of effort are expected to lead to high levels of performance, (b) *instrumentality,* or the degree to which high-quality performance is perceived as instrumental in obtaining an outcome, and (c) *valence* of the outcome, or the degree to which the outcome is viewed as desirable. The CEM expands on this logic by specifying that instrumentality on collective tasks is determined by three factors: (a) the perceived relationship between individual performance and group performance, (b) the perceived relationship between group performance and group outcomes, and (c) the perceived relationship between group outcomes and individual outcomes. Thus, the model suggests that working on a collective task introduces additional contingencies between individuals' efforts and their outcomes. For this reason, an analysis of individual effort is by necessity more complex for collective tasks than for coactive tasks.

The CEM also acknowledges that, although individual and group performance are strongly related to many of the valued outcomes in performance settings, there are some valued outcomes that do not depend on performance. For example, when working on intrinsically meaningful tasks or when working with highly respected co-workers, exerting high levels of effort may lead to self-satisfaction, approval from the group, or other important outcomes, even if such high effort has little or no impact on tangible performance outcomes. However, the model emphasizes performance-related outcomes because many of the most important outcomes available in a performance context are contingent on actual performance.

The CEM can be described as "cognitive" both because perceived rather than actual contingencies are hypothesized to influence behavior and because individuals are hypothesized to either consciously or subconsciously select a level of effort to exert on the task. However, the model is not necessarily deliberative because individuals are unlikely to systematically process all of the available information about the task performance situation, unless they are particularly motivated to process it because of situational constraints or individual differences. Therefore, some situations may lead individuals to respond automatically to a preexisting effort script, whereas other situations may lead individuals to strategically increase or decrease their collective effort. The CEM suggests that individuals generally consider salient features of the task setting when attempting to maximize the expected utility of their actions.

The CEM also expands on prior expectancy-value frameworks by specifying which outcomes are likely to be valued in collective settings. Valued outcomes can consist of either objective outcomes such as pay or subjective outcomes such as enjoyment, satisfaction, feelings of group spirit and belonging, and feelings of self-worth. In the case of objective outcomes, however, it is the individual's evaluation of the outcome rather than the outcome itself that determines its valence (e.g., Deci, 1975; Lepper & Greene, 1978). Because the CEM is oriented to explaining collective phenomena, the model places particular emphasis on group-level outcomes that have implications for the individual's self-evaluation. Research has repeatedly demonstrated that individuals are quite concerned with maintaining a

favorable self-evaluation (e.g., Brown, Collins, & Schmidt, 1988; Greenwald, 1982; Greenwald & Pratkanis, 1984; Suls & Miller, 1977; Tesser, 1988). Group performance situations produce the potential for self-evaluation from a variety of relevant sources (Crocker & Luhtanen, 1990; Leary & Forsyth, 1987). Breckler and Greenwald (1986) proposed that the self has distinct motivational facets that are influenced most strongly by three corresponding audiences. They distinguished between (a) a private self that responds primarily to internal values, (b) a public self that responds mainly to others' evaluations, and (c) a collective self that is sensitive primarily toward fulfilling one's role within the context of the goals of important reference groups. Although most social loafing research has been focused on the motivation produced by the experimenter's evaluation of performance (for prominent exceptions, see Harkins & Szymanski, 1988, 1989), Breckler and Greenwald's analysis suggests that individuals can be motivated by appeals to any of the three motivational facets of the self (see also Crocker & Luhtanen, 1990). With respect to the CEM, task situations that provide clear information relevant to self-evaluation, whether from oneself, one's co-workers, reference groups, or others, should be more inherently motivating than situations that do not provide such information or that make such information ambiguous and less potent.

Social identity theory and group-level versions of social comparison theory lead to similar conclusions. Social identity theory (e.g., Abrams & Hogg, 1990; Tajfel & Turner, 1986) proposes that individuals gain positive self-identity through the accomplishments of the groups to which they belong (see also Cialdini et al., 1976; Levine & Moreland, 1987). Indeed, several recent analyses suggest that some social motivations, such as a need for belonging or a need for social communication, can only be fulfilled in a collective setting (e.g., Brewer, 1991; Caporael, Dawes, Orbell, & van de Kragt, 1989). Similarly, Goethals and Darley's (1987) group-level revision of social comparison theory suggests that individuals not only compare themselves with other individuals to obtain self-evaluation information, but that they also compare the groups to which they belong to other groups for the same reason. Goethals and Darley further stated that individuals may seek such group-level comparison information both to obtain self-knowledge (information about the actual level of one's performance compared with others) and self-validation (information that level of one's performance is superior to that of others). Taken as a whole, the recent theories of group level self-evaluation processes all suggest that collective contexts that provide a great deal of information relevant to one's self-validation should be more important to individuals than contexts that provide less information (or none). These perspectives also suggest that self-evaluation information can come from a variety of sources and that information relevant to one's role in valued reference groups may be especially influential.

Predictions for Meta-Analysis

On the basis of the preceding logic, the CEM provides a number of predictions for our meta-analysis. First, because the model suggests that individuals' outcomes are frequently less reliant on their efforts when working collectively than coactively, we predicted that individuals would generally tend to engage in social loafing and that this effect would be reliable across studies. Second, the CEM suggests that individuals will work harder on a collective task when they expect their efforts to be instrumental in obtaining valued outcomes. Therefore, we predicted that, holding other factors constant, social loafing should be reduced when individuals: (a) believe that their collective performances can be evaluated by the experimenter, their co-workers, themselves, or others; (b) work in smaller rather than larger groups; (c) perceive that their contributions to the collective product are unique, rather than redundant with the inputs of others; (d) are provided with a standard with which to compare their group's performance; (e) work on tasks that are either intrinsically interesting, meaningful to the individual, important to one's reference group or to valued others, or high in personal involvement; (f) work with respected others (high group valence; friends, teammates, partners, and respected co-workers) or in a situation that activates a salient group identity; (g) expect their co-workers to perform poorly; and (h) have a dispositional tendency to view favorable collective outcomes as valuable and important.

Regarding the final prediction, several factors are likely to influence the degree to which an individual values collective outcomes relative to individual outcomes. Two especially prominent factors may be gender and culture. Recent gender research seems to suggest that women tend to be more group- or collectively-oriented than men. For example, research on gender stereotypes (e.g., Broverman, Vogel, Broverman, Clarkson, & Rosenkrantz, 1972; Eagly & Steffen, 1984) has demonstrated that men are expected to possess high levels of agentic qualities (such as being independent and assertive), whereas women are expected to possess high levels of communal qualities (such as being friendly, unselfish, and concerned with others; see Bakan, 1966). Similarly, research conducted in small-group settings has found that men tend to specialize in activity oriented toward task completion, whereas women tend to specialize in activity oriented toward meeting interpersonal demands within the group (Anderson & Blanchard, 1982; Carli, 1982). Taken together, these findings suggest that men appear to be oriented toward individualistic and competitive concerns, whereas women appear to be oriented toward interpersonal and cooperative concerns. An extension of this logic to social loafing suggests that women are more likely than men to view performing well on collective tasks as important. Thus, we predicted that, although both sexes are likely to engage in social loafing, women should tend to loaf less than men.

In a similar manner, the degree to which the dominant culture from which subject populations were selected emphasizes individualistic versus collectivistic concerns may moderate the social loafing effect. Eastern or Oriental culture is often depicted as group- or socially oriented, whereas Western or American culture is often depicted as individualistically oriented (e.g., Hsu, 1970; Triandis, 1989; Wheeler, Reis, & Bond, 1989). Therefore, subjects from Eastern cultures are more likely to view performance on collective tasks as important than are subjects from Western cultures. Thus, we predicted that subjects in countries such as Japan, Taiwan, and China should loaf less than subjects in the United States and Canada.

Finally, as discussed earlier, Jackson and Williams (1985)

proposed that the presence of others is drive reducing when other group members serve as cotargets of an outside source of social influence. According to drive models of anxiety, working collectively should lead to reduced drive, resulting in decreased performance on simple tasks and increased performance on novel, difficult, tasks. In our meta-analysis, we addressed the adequacy of this viewpoint by classifying studies according to whether they used simple or complex tasks.

Some of our predictions mirror well-established effects in the social-loafing literature, whereas other predictions have either been examined only sporadically, if at all, or have produced conflicting evidence. For example, researchers generally accept the conclusion that individuals tend to engage in social loafing when working collectively, and a number of studies have found that social loafing can be reduced or eliminated when evaluation is provided in the collective condition. However, the effects of group-level comparison standards, task meaningfulness, and task uniqueness have only been examined in a handful of studies. The degree to which these variables consistently influence social loafing across studies is not known, and only tentative conclusions about their effects can currently be drawn. Finally, contradictory results have emerged from studies examining gender, culture, expectations of co-worker performance, group size, and group valence. In the present review, we quantitatively examine the influence of all of these variables on social loafing, using all available comparisons, to provide a clearer understanding of what factors moderate social loafing in different situations. With respect to well-established findings, our meta-analytic methods allowed us to estimate the magnitude of social loafing and examine its consistency across studies. Similarly, we were able to assess quantitatively the impact of evaluation on social loafing and compare it with the influence of other variables affecting collective effort.

Method

Sample of Studies

Computer-based information searches were conducted using the key words *social loafing, free-rider, Ringelmann, motivation loss, motivation decrement, collective task,* and *collective performance.* These key words were searched in the following databases: *Psychological Abstracts* (PsycINFO), *Dissertation Abstracts International* (DISS), *Sociological Abstracts* (SOCA), Educational Resources Information Center (ERIC), and a worldwide business and management database (ABI/INFORM). In addition, the reference lists of numerous review articles and chapters, as well as the reference lists of all located studies, were searched. Finally, several researchers provided data from unpublished studies that had been presented at regional or national meetings.

Three key criteria were adopted for including studies in the sample. First, the study had to compare the effort of individuals working either individually or coactively with the effort of individuals working collectively. Collective performance was defined as a situation in which individuals' inputs were combined into one group score. Thus, studies were excluded if they merely compared the effort of individuals across different group sizes without including a coactive control condition (e.g., Kerr & Bruun, 1983, Experiments 1 and 2; Zaccaro, 1984): The individual-versus-group work context, which is central to the social loafing construct, is not manipulated in such studies. This criterion also excluded studies in which only evaluation potential was manipulated, without providing subjects with an expectation that their inputs would be combined with those of other subjects (e.g., Price, 1987, Experiment 1).

Second, studies were excluded if the number of task performers (i.e., coactors or co-workers) present was confounded across coactive and collective conditions (e.g., Huddleston, Doody, & Ruder, 1985; Kerr & Bruun, 1981, Experiment 1, between-subjects comparisons). In such studies, effort cannot be separated from distraction and mere presence effects. Of course, this criterion excluded most social facilitation studies, because they intentionally varied the number of individuals present in the alone and group conditions.[2]

Studies were also omitted if individual effort could not be separated from coordination loss (e.g., Ingham et al., 1974, Experiment 1; see introduction). This criterion excluded studies that merely compared the average individual performance with the average group performance without measuring individual performances under collective conditions.

Variables Coded From Each Study

The following general information was coded for each study: (a) date of publication, (b) publication form (journal article, meeting paper, dissertation or master's thesis, or unpublished document), (c) number of subjects, (d) age of subjects, and (e) status of subjects (third grade students or younger, fourth through sixth grade students, junior high or high school students, college students, or organizational employees). The following methodological characteristics were also coded: (a) setting of study (laboratory or field), (b) type of coactive–collective comparison (within-subjects or between-subjects), and (c) type of cover story used (effort or performance related or other, e.g., studying the effects of "sensory sound dynamics," improving writing in a college journalism department, or obtaining evaluations of clinical therapists).

In addition, the following attributes of the task were coded: (a) complexity of task (simple or well-learned, complex or novel, or unknown or unclear); (b) type of effort required by task (physical, e.g., shouting or rope-pulling; cognitive, e.g., generating ideas; perceptual, e.g., vigilance, maze performance; or evaluative, e.g., quality rating of poems or editorials); (c) quantity or quality emphasis of task using Steiner's (1972) typology (maximizing, optimizing, and mixed or unclear); and (d) method of combining individual inputs, again using Steiner's typology (additive, compensatory, disjunctive, and mixed or unclear).

Finally, the following predictors relevant to the CEM were also coded: (a) the conditions under which individual inputs could be evaluated, that is, evaluation potential (none, coactive condition only, coactive and collective conditions, and unclear); (b) task valence (high, low, and unclear or unspecified); (c) group valence (high, e.g., close friends or couples, group cohesiveness manipulation, or teammates; moderate, e.g., mere acquaintances; low, e.g., strangers; unknown or unclear, e.g., clear possibility that subjects might be acquainted but not stated in report); (d) opportunity for group evaluation (present or absent); (e) expectations of co-worker performance (high, low, or unclear or unspecified); (f) uniqueness of individual task contributions (unique, e.g., brainstorming tasks in which each subject generates uses for a different

[2] Several studies (e.g., Harkins & Szymanksi, 1988; Weldon & Gargano, 1985) had subjects perform a task "alone" in the collective condition by leading these individuals to believe that their inputs would be combined with those of others who would participate at different sessions, thus holding group size constant across individual and collective conditions. There were no significant differences between studies that had subjects work alone (across individual and collective conditions) and studies that had subjects work in the presence of others (across coactive and collective conditions).

object or vigilance tasks in which each subject watches for randomly occurring blips in different quadrants of a video terminal; potentially redundant, e.g., brainstorming tasks in which all subjects generate uses for the same object, evaluative tasks that use a compensatory method of combining inputs; completely redundant, e.g., pulling on a rope, shouting and clapping, or pumping air); (g) group size (i.e., number of individual inputs combined into the group product); (h) number of task performers present at session; (i) sex of subjects (men, women, or both); and (j) culture of subjects (Eastern, i.e., China, Japan, or Taiwan or Western, i.e., United States or Canada).

Participants' efforts were only coded as evaluable if (a) their output was identifiable either to the experimenter, their co-workers, or themselves, and (b) a personal, social, or objective standard was available with which their outputs could be compared. This treatment is consistent with widely accepted interpretations of evaluation potential (e.g., Harkins, 1987; Harkins & Szymanski, 1988, 1989).

Task valence was coded as high (or low) if researchers manipulated task importance, task meaningfulness, or personal involvement in such a way that subjects were provided with a clear rationale for viewing the task as especially important or trivial (e.g., Brickner, Harkins, & Ostrom, 1986; Williams & Karau, 1991). Task valence was also coded as high (or low) if subjects were selected on a personality variable that would make them likely to view the task used as especially involving (or aversive), or were selected on the basis of their probable intrinsic interest (or disinterest) in the task (e.g., subjects high and low in need for cognition performing an idea-generation task—Petty, Cacioppo, & Kasmer, 1985; collegiate swimmers at a competitive meet—Williams, Nida, Baca, & Latané, 1989). In all other cases, task valence was scored as unclear or unspecified. For practical purposes, tasks that were scored as unclear or unspecified are likely to fall somewhere between high- and low-valence tasks, and therefore can reasonably be regarded as somewhat moderate in valence.

Expectations of co-worker performance were coded as high (or low) on the basis of researchers' manipulations of co-worker ability or co-worker effort, as well as on the basis of subjects' selection on interpersonal trust scores (low trusters may expect others to engage in social loafing, cf. Williams & Karau, 1991). Expectations of co-worker performance were also coded as high if subjects were experts at the task (e.g., collegiate swimmers). In all other cases, expectations were coded as unclear or unspecified.

The categories listed for some of these variables were simplified from an initially more detailed set of categories either because (a) some categories had few or no entries or (b) prediction of the effect sizes was not improved by the more detailed categories. The variables for which the initial coding was more detailed are the following: status of subjects, culture of subjects, type of effort measured, method of combining individual inputs, and evaluation potential.

All variables were coded by Steven J. Karau. To estimate reliability, Kipling D. Williams independently coded a subsample of 19 studies (24%). The median agreement was 100% (M agreement = 98%, kappa = .95). "Complexity of task" yielded the lowest agreement, 83% (kappa = .71). Disagreements were resolved by discussion. A complete list of the studies included in the meta-analysis, along with their characteristics and effect sizes, is presented in Table 1.

Computation and Analysis of Effect Sizes

The effect size calculated is g, the difference between individual effort in the coactive and collective conditions, divided by the pooled standard deviation (see Hedges & Olkin, 1985). A positive sign indicates that subjects exerted more effort coactively than collectively (i.e., engaged in social loafing), and a negative sign indicates that subjects exerted more effort collectively than coactively. The effect sizes were calculated by Steven J. Karau with the aid of a computer program (Johnson, 1989). Although objective effort measures are of central concern to this meta-analysis, effect sizes based on self-report measures of effort were also calculated so that supplementary analyses could be conducted. Several studies provided data for more than one relevant dependent variable. When this occurred, separate effect sizes were computed for each dependent variable, and these separate effect sizes were then combined using Rosenthal and Rubin's (1986) suggested formula.[3]

Multiple effect sizes from single studies. Under some circumstances, studies were partitioned and separate effect sizes were computed within levels of an independent variable. This partitioning was undertaken when (a) subjects of different status or from different cultures were used (7 studies); (b) group size varied (11 studies); (c) different tasks were used that varied either in their complexity, the uniqueness of individual inputs to the collective product, or the method used to combine individual inputs (10 studies); (d) evaluation potential varied (17 studies); (e) task valence varied (6 studies); (f) group valence differed (5 studies); or (g) expectations of co-worker performance were manipulated (7 studies). Although this partitioning of studies created some nonindependence in the data, the strategy was necessary to allow theory-relevant interactions in the studies to be represented and to test the CEM's predictions about moderating variables. A study that manipulated one of these variables was partitioned only if its reported findings were sufficient to allow the computation of effect sizes within the levels of the variable or to determine the direction or significance of findings within the levels. By this strategy, the original sample of 78 studies obtained from 59 documents produced 166 units, 26 of which were intact studies and 140 of which were subdivided parts of studies. For simplicity of exposition, these units are referred to as *studies* in the remainder of this manuscript. For some analyses, the 166 units were further partitioned by sex of subject when data were separately reported for male and female subjects. When this latter partitioning was undertaken, 178 units were available for analysis.

Analysis of effect sizes. The analysis was based on the methods detailed by Hedges and Olkin (1985). The gs were converted to ds by correcting them for bias. To obtain an overall estimate of the difference in individuals' coactive and collective effort reported in the available research, the relevant study outcomes were then combined by averaging the ds. All such means were computed with each effect size weighted by the reciprocal of its variance, a procedure that gives more weight to effect sizes that are more reliably estimated.[4] A homogeneity statistic, Q, was also calculated to determine whether each set of ds shared a common effect size (i.e., was consistent across the studies). To test the CEM's predictions about moderating variables, we accounted for variability in the effect sizes by relating them to attributes of the studies using both categorical and continuous models. The categorical models, which are analogous to analyses of variance (ANOVAs), provide a between-classes effect, Q_B (analogous to a main effect in an

(*text continues on page 694*)

[3] The between-measures correlations that were used when implementing this formula were estimated from correlations (a) given in several studies in the sample and (b) estimated from data provided by Williams (from Williams & Williams, 1981, Experiment 1) and Gabrenya (from Gabrenya, Latané, & Wang, 1983). The resulting correlations were .82 for subjective effort, .71 for objective measures based on shouting and clapping tasks, and .49 for objective measures based on evaluative tasks.

[4] Because some studies used the group as the unit of analysis, whereas others used individuals or individuals across repeated factors as the unit of analysis, the number of subjects in the coactive and collective conditions were used in all cases to represent sample size in these calculations in an attempt to avoid arbitrarily weighting some effect sizes dramatically higher than others despite comparable numbers of subjects.

Table 1
Studies and Their Characteristics

Study and subset	Task	Effect size (*d*)	Subjects[a]	Methodology[b]	Task[c]	Theory-relevant predictors[d]	
Bartis, Szymanski, & Harkins (1988)	Brainstorming						
Maximizing instructions		0.63	52/20.00/4	1/1/2/2	1/2/1/1	2/2/3/2/2	2/—/1.00/3/2
Optimizing instructions		0.11	51/20.00/4	1/3/2/2	2/2/2/1	2/2/3/2/2	2/—/1.00/3/2
Brickner, Harkins, & Ostrom (1986)	Thought listing						
High personal involvement		0.13	112/20.00/4	1/2/2/2	3/4/3/1	2/1/3/2/2	2/2.00/2.00/3/2
Low personal involvement		0.88	112/20.00/4	1/1/2/2	3/4/3/1	2/2/3/2/2	2/2.00/2.00/3/2
Brickner & Wingard (1988)	Finding items in picutre						
High group identity		−0.06	30/20.00/4	1/2/2/1	1/3/1/1	2/2/1/2/2	2/—/2.00/3/2
Low group identity		0.89	30/20.00/4	1/1/2/1	1/3/1/1	2/2/3/2/2	2/—/2.00/3/2
Cates (1986)	Brainstorming						
Dissertation							
Unique object, high identifiability		−0.05	40/20.00/4	1/2/2/1	1/2/1/1	3/2/3/2/2	1/4.00/4.00/3/2
Unique object, low identifiability		0.05	40/20.00/4	1/2/2/1	1/2/1/1	2/2/3/2/2	1/4.00/4.00/3/2
Common object, high identifiability		0.36	40/20.00/4	1/2/2/1	1/2/1/1	3/2/3/2/2	2/4.00/4.00/3/2
Common object, low identifiability		1.20	40/20.00/4	1/1/2/1	1/2/1/1	2/2/3/2/2	2/4.00/4.00/3/2
Pilot study		0.79	37/20.00/4	1/1/2/1	1/2/1/1	2/2/3/2/2	2/4.00/18.67/3/2
Cohen (1988)	Shouting and clapping	0.12	312/20.00/4	1/1/1/2	1/1/1/1	2/2/3/2/2	3/2.00/2.00/3/2
Cooley (1991)	Brainstorming						
Student sample		−0.04	88/20.14/4	1/2/2/1	1/2/1/1	2/1/3/2/2	2/5.00/5.00/3/2
Adult sample		−0.32	33/38.70/5	1/2/2/1	1/2/1/1	2/1/4/2/2	2/5.00/5.00/3/2
Earley (1989)	In-basket						
High accountability, American sample		0.29	24/23.82/5	2/2/2/1	3/2/1/1	3/1/2/2/2	2/10.00/—/3/2
High accountability, Chinese sample		−2.32	24/24.33/5	2/2/2/1	3/2/1/1	3/1/2/2/2	2/10.00/—/3/1
Low accountability, American sample		3.19	24/23.82/5	2/1/2/1	3/2/1/1	4/1/2/2/2	2/10.00/—/3/2
Low accountability, Chinese sample		−2.95	24/24.33/5	2/2/2/1	3/2/1/1	4/1/2/2/2	2/10.00/—/3/1
Gabrenya, Latané, & Wang (1981)	Shouting and clapping						
American undergraduates		0.73	36/20.00/4	1/1/1/2	1/1/1/1	2/2/4/2/2	3/2.00/2.00/1/2
American graduate students		0.20	32/24.00/4	1/1/1/2	1/1/1/1	2/2/4/2/2	3/2.00/2.00/1/2
Chinese graduate students		−0.39	34/24.00/4	1/2/1/2	1/1/1/1	2/2/4/2/2	3/2.00/2.00/1/1
Gabrenya, Latané, & Wang (1983)	Shouting and clapping						
Main experiment							
Second and third graders		0.53	80/ 7.50/1	2/2/1/1	1/1/1/1	2/2/2/2/2	3/2.00/2.00/3/1
Sixth graders		0.72	40/11.00/2	2/2/1/1	1/1/1/1	2/2/2/2/2	3/2.00/2.00/3/1
Ninth graders		0.89	40/14.00/3	2/2/1/1	1/1/1/1	2/2/2/2/2	3/2.00/2.00/3/1
Study cited in footnote		0.74	19/ 5.00/1	2/2/1/1	1/1/1/1	2/2/4/2/2	3/2.00/2.00/3/1
Gabrenya, Wang & Latané (1985)	Counting tones						
Main experiment							
Effort trials, American sixth graders		0.04	32/11.00/2	2/1/1/1	3/3/2/1	4/2/4/2/2	3/2.00/2.00/3/2
Effort trials, American ninth graders		0.23	40/14.00/3	2/1/1/1	3/3/2/1	4/2/4/2/2	3/2.00/2.00/3/2
Effort trials, Chinese sixth graders		0.13	40/11.00/2	2/2/1/1	3/3/2/1	4/2/4/2/2	3/2.00/2.00/3/1
Effort trials, Chinese ninth graders		−0.11	40/14.00/3	2/2/1/1	3/3/2/1	4/2/4/2/2	3/2.00/2.00/3/1
Choice trials		n.s.	152/12.58/1	2/2/1/1	3/3/2/1	3/2/4/2/2	3/2.00/2.00/3/3
Study cited in footnote		0.34	58/20.00/4	1/1/1/2	3/3/2/1	4/2/3/2/2	3/2.00/2.00/1/2
Griffith, Fichman, & Moreland (1989)	Maze performance						
Alone, simple task		−0.31	16/20.00/4	1/1/2/2	1/3/3/1	2/2/3/2/2	3/—/1.00/3/2
Alone, complex task		−0.17	16/20.00/4	1/2/2/2	2/3/3/1	2/2/3/2/2	3/—/1.00/3/2
Group, simple task		0.93	16/20.00/4	1/1/2/2	1/3/3/1	2/2/3/2/2	3/—/2.00/3/2
Group, complex task		0.47	16/20.00/4	1/2/2/2	2/3/3/1	2/2/3/2/2	3/—/2.00/3/2
Hallmark & Downs (1987)	Brainstorming	0.88	31/20.00/4	2/1/1/2	1/2/1/1	4/2/4/2/2	2/—/—/3/2
Harcum & Badura (1990)							
Perceptual task	Word circling	0.22	40/20.00/4	1/1/2/1	1/3/3/1	2/2/3/2/2	3/2.00/1.00/3/2
Physical task	Tapping task	1.01	40/20.00/4	1/1/2/1	1/1/1/1	2/2/3/2/2	3/2.00/1.00/3/2

Table 1 (*continued*)

Study and subset	Task	Effect size (*d*)	Subjects[a]	Methodology[b]	Task[c]	Theory-relevant predictors[d]	
Hardy & Latané (1986), Experiment 1	Shouting and clapping	0.41	48/20.00/4	1/1/1/2	1/1/1/1	2/2/3/2/2	3/2.00/2.00/3/2
Hardy & Latané (1988)	Shouting and clapping						
Cohesive groups		0.13	24/16.00/3	1/2/1/1	1/1/1/1	2/1/1/2/2	3/2.00/2.00/2/2
Noncohesive groups		0.55	24/16.00/3	1/1/1/1	1/1/1/1	2/1/3/2/2	3/2.00/2.00/2/2
Harkins (1987), Experiment 1	Brainstorming						
Evaluation condition		0.92	48/20.00/4	1/1/2/1	1/2/1/1	2/2/3/2/2	2/2.00/2.00/3/2
No evaluation condition		−0.17	48/20.00/4	1/2/2/1	1/2/1/1	1/2/3/2/2	1/2.00/2.00/3/2
Harkins & Jackson (1985)	Brainstorming						
Comparable uses condition		0.77	80/20.00/4	1/1/2/1	1/2/1/1	2/2/3/2/2	2/4.00/4.00/3/2
Non-comparable uses condition		0.06	80/20.00/4	1/2/2/1	1/2/1/1	1/2/3/2/2	1/4.00/4.00/3/2
Harkins, Latané, & Williams (1980)	Clapping						
Experiment 1							
Within-subjects comparisons		1.30	48/20.00/4	1/1/1/2	1/1/1/1	2/2/3/2/2	3/2.00/4.00/3/2
Between-subjects comparisons		1.05	48/20.00/4	1/1/2/2	1/1/1/1	2/2/3/2/2	3/2.00/2.00/3/2
Experiment 1 replication		1.49	48/20.00/4	1/1/2/2	1/1/1/1	2/2/3/2/2	3/2.00/2.00/3/2
Harkins & Petty (1982)							
Experiment 1	Brainstorming						
Easy task		0.83	32/20.00/4	1/1/2/1	1/2/1/1	2/2/3/2/2	2/10.00/3.00/3/2
Difficult task		−0.25	32/20.00/4	1/2/2/1	2/2/1/1	2/2/3/2/2	1/10.00/3.00/3/2
Experiment 1 replication	Brainstorming						
Easy task		1.22	28/20.00/4	1/1/2/1	1/2/1/1	2/2/3/2/2	2/10.00/3.00/3/2
Difficult task		−0.07	28/20.00/4	1/2/2/1	2/2/1/1	2/2/3/2/2	1/10.00/3.00/3/2
Experiment 2	Vigilance						
Easy task		2.28	61/20.00/4	1/1/2/1	1/3/2/1	2/2/3/2/2	3/3.81/3.81/3/2
Difficult task		−0.17	61/20.00/4	1/2/2/1	2/3/2/1	2/2/3/2/2	3/3.81/3.81/3/2
Experiment 3	Vigilance						
Unique task		0.43	43/20.00/4	1/2/2/1	3/3/2/1	2/2/3/2/2	1/3.58/3.58/3/2
Redundant task		2.00	43/20.00/4	1/1/2/1	3/3/2/1	2/2/3/2/2	3/3.58/3.58/3/2
Harkins & Szymanski (1988)	Vigilance						
Output feedback, standard		0.13	30/20.00/4	1/2/2/1	1/3/2/1	3/2/3/2/2	3/—/1.00/3/2
Output feedback, no standard		1.39	30/20.00/4	1/1/2/1	1/3/2/1	2/2/3/2/2	3/—/1.00/3/2
No output feedback, standard		0.30	30/20.00/4	1/1/2/1	1/3/2/1	3/2/3/2/2	3/—/1.00/3/2
No output feedback, no standard		1.67	30/20.00/4	1/1/2/1	1/3/2/1	2/2/3/2/2	3/—/1.00/3/2
Harkins & Szymanski (1989)							
Experiment 1	Vigilance						
Standard		−0.13	72/20.00/4	1/2/2/1	1/3/2/1	2/2/3/1/2	3/3.00/3.00/3/2
No standard		1.85	72/20.00/4	1/1/2/1	1/3/2/1	2/2/3/2/2	3/3.00/3.00/3/2
Experiment 2	Brainstorming						
Standard, feedback		−0.22	96/20.00/4	1/2/2/1	1/2/1/1	2/2/3/1/2	2/3.00/3.00/3/2
No standard, feedback		1.17	96/20.00/4	1/1/2/1	1/2/1/1	2/2/3/2/2	2/3.00/3.00/3/2
No standard, no feedback		1.35	96/20.00/4	1/1/2/1	1/2/1/1	2/2/3/2/2	2/3.00/3.00/3/2
Ingham, Levinger, Graves, & Peckham (1974), Experiment 2	Rope-pulling						
Group size = 2		0.58	36/20.00/4	1/1/1/1	1/1/1/1	2/2/3/2/2	3/2.00/6.00/1/2
Group size = 3		0.92	36/20.00/4	1/1/1/1	1/1/1/1	2/2/3/2/2	3/3.00/6.00/1/2
Group size = 4		0.77	36/20.00/4	1/1/1/1	1/1/1/1	2/2/3/2/2	3/4.00/6.00/1/2
Group size = 5		1.10	36/20.00/4	1/1/1/1	1/1/1/1	2/2/3/2/2	3/5.00/6.00/1/2
Group size = 6		0.86	36/20.00/4	1/1/1/1	1/1/1/1	2/2/3/2/2	3/6.00/6.00/1/2
Jackson & Harkins (1985)	Shouting						
High co-worker effort		0.15	14/20.00/4	1/2/1/1	1/1/1/1	2/1/3/2/1	3/2.00/2.00/2/2
Low co-worker effort		0.37	14/20.00/4	1/2/1/1	1/1/1/1	2/3/3/2/3	3/2.00/2.00/2/2
Social loafing replication		1.31	14/20.00/4	1/1/1/1	1/1/1/1	2/2/3/2/2	3/2.00/2.00/2/2
Jackson & Latané (1982a)	Shouting	0.88	64/20.00/4	1/1/1/2	1/1/1/1	2/2/3/2/2	3/2.00/2.00/3/2
Jackson & Latané (1982b)	Shouting	0.31	80/20.00/4	1/1/1/2	1/1/1/1	2/2/3/2/2	3/4.00/4.00/3/2

Table 1 (*continued*)

Study and subset	Task	Effect size (*d*)	Subjects[a]	Methodology[b]	Task[c]	Theory-relevant predictors[d]	
Jackson & Williams (1985)	Maze performance						
Simple task		0.33	32/20.00/4	1/1/2/1	1/3/3/1	2/2/3/2/2	2/2.00/2.00/3/2
Complex task		−0.52	32/20.00/4	1/2/2/1	2/3/3/1	2/2/3/2/2	2/2.00/2.00/3/2
Jackson, Williams, & Latané (1978)	Clapping						
Group size = 2		0.32	96/20.00/4	1/1/1/2	1/1/1/1	2/2/3/2/2	3/2.00/4.00/3/2
Group size = 4		0.43	96/20.00/4	1/1/1/2	1/1/1/1	2/2/3/2/2	3/4.00/4.00/3/2
Karau & Williams (1993), Experiment 2	Brainstorming						
Cohesive groups, high co-worker ability		−0.40	28/20.00/4	1/2/2/1	1/2/1/1	2/1/1/2/1	2/2.00/2.00/3/2
Cohesive groups, low co-worker ability		−0.41	29/20.00/4	1/2/2/1	1/2/1/1	2/1/1/2/3	2/2.00/2.00/3/2
Cohesive groups, no mention of ability		−0.12	30/20.00/4	1/2/2/1	1/2/1/1	2/1/1/2/2	2/2.00/2.00/3/2
Noncohesive groups, high co-worker ability		0.58	26/20.00/4	1/1/2/1	1/2/1/1	2/1/3/2/1	2/2.00/2.00/3/2
Noncohesive groups, low co-worker ability		−0.84	26/20.00/4	1/2/2/1	1/2/1/1	2/1/3/2/3	2/2.00/2.00/3/2
Noncohesive groups, no mention of ability		0.88	25/20.00/4	1/1/2/1	1/2/1/1	2/1/3/2/2	2/2.00/2.00/3/2
Kerr (1983)	Air pumping						
High co-worker effort		1.22	29/20.00/4	1/1/2/1	1/1/1/3	3/2/3/1/1	3/2.00/2.00/3/2
Low co-worker effort		1.17	31/20.00/4	1/1/2/1	1/1/1/3	3/2/3/1/3	3/2.00/2.00/3/2
Low co-worker ability		0.64	30/20.00/4	1/2/2/1	1/1/1/3	3/2/3/1/3	3/2.00/2.00/3/2
Kerr & Bruun (1981)							
Experiment 1, within-subjects comparisons							
Group size = 2		0.75	40/20.00/4	1/1/1/1	1/1/1/1	2/2/3/2/2	3/2.00/5.00/3/2
Group size = 4		1.09	40/20.00/4	1/1/1/1	1/1/1/1	2/2/3/2/2	3/4.00/5.00/3/2
Experiment 2							
Group size = 2		0.31	28/20.00/4	1/3/1/1	1/1/1/1	4/2/3/2/2	3/2.00/5.00/3/2
Group size = 4		0.24	28/20.00/4	1/3/1/1	1/1/1/1	4/2/3/2/2	3/4.00/5.00/3/2
Kerr & Bruun (1983), Experiment 3	Air pumping						
No feedback, group size = 2		0.04	108/20.00/4	1/3/1/1	1/1/1/3	3/2/3/2/2	3/2.00/3.00/2/2
No feedback, group size = 3		0.15	108/20.00/4	1/3/1/1	1/1/1/3	3/2/3/2/2	3/3.00/3.00/2/2
Feedback, group size = 2		0.12	108/20.00/4	1/3/1/1	1/1/1/3	3/2/3/1/2	3/2.00/3.00/2/2
Feedback, group size = 3		0.22	108/20.00/4	1/3/1/1	1/1/1/3	3/2/3/1/2	3/3.00/3.00/2/2
Latané, Williams, & Harkins (1979), Experiment 1	Shouting						
Group size = 2		1.34	36/20.00/4	1/1/1/2	1/1/1/1	2/2/3/2/2	3/2.00/6.00/1/2
Group size = 6		1.94	36/20.00/4	1/1/1/2	1/1/1/1	2/2/3/2/2	3/6.00/6.00/1/2
Martin, Seta, & Crelia (1990), Experiment 2	Impression formation	1.05	62/20.00/4	1/1/2/2	1/4/2/2	2/2/3/2/2	2/4.50/3.00/3/2
Matsui, Kakuyama, & Onglatco (1987), Experiment 1	Perceptual counting task	−0.63	104/20.00/4	1/2/2/1	1/3/1/1	3/2/3/1/2	3/2.00/—/3/1
McGovern (1986)	Vigilance	0.85	64/20.00/4	1/3/2/1	1/3/2/1	2/2/3/2/2	3/4.00/4.00/3/2
Peterson, Zaccaro, & Daly (1986)	Brainstorming	2.04	41/20.00/4	1/1/2/1	1/2/1/1	2/2/3/2/2	2/20.00/20.00/3/2
Petty, Cacioppo, & Kasmer (1985)	Brainstorming						
High need for cognition		0.03	40/20.00/4	1/2/2/1	1/2/1/1	2/1/3/2/2	2/6.50/6.50/3/2
Low need for cognition		0.94	40/20.00/4	1/1/2/1	1/2/1/1	2/3/3/2/2	2/6.50/6.50/3/2
Petty, Harkins, & Williams (1980)							
Experiment 1	Evaluating therpaists	0.59	52/20.00/4	1/1/2/2	3/4/2/3	2/2/3/2/2	2/15.00/1.00/3/2
Experiment 2	Evaluating messages	0.84	180/20.00/4	1/1/2/2	3/4/2/2	2/2/3/2/2	2/10.00/11.00/3/2
Petty, Harkins, Williams, & Latané (1977)	Evaluating written materials						
Group size = 4		—	50/20.00/4	1/1/2/2	3/4/2/2	2/2/3/2/2	2/4.00/—/3/2
Group size = 16		—	50/20.00/4	1/1/2/2	3/4/2/2	2/2/3/2/2	2/16.00/—/3/2
Price (1987), Experiment 2	Evaluating information						
High identifiability		0.08	94/20.00/4	1/2/2/2	3/4/2/2	3/1/3/2/2	2/10.00/10.00/3/2
Low identifiability		−0.61	94/20.00/4	1/1/2/2	3/4/2/2	1/1/3/2/2	2/10.00/10.00/3/2
Quintanar & Pryor (1982)	Making causal attributions	0.70	36/20.00/4	1/1/2/2	1/4/2/2	2/2/3/2/2	2/10.00/7.00/3/2
Shepperd & Wright (1989)	Brainstorming						
Incentive		−0.04	40/20.00/4	1/2/2/1	1/2/1/1	2/1/3/2/2	2/4.50/4.50/3/2
No incentive		0.81	40/20.00/4	1/1/2/1	1/2/1/1	2/2/3/2/2	2/4.50/4.50/3/2

Table 1 (*continued*)

Study and subset	Task	Effect size (*d*)	Subjects[a]	Methodology[b]	Task[c]	Theory-relevant predictors[d]	
Shirakashi (1985)	Shouting and clapping						
Cohesive groups		0.01	32/19.50/4	1/2/1/2	1/1/1/1	2/2/1/2/2	3/2.00/2.00/1/1
Noncohesive group		0.23	32/21.00/4	1/3/1/2	1/1/1/1	2/2/3/2/2	3/2.00/2.00/1/1
Sorrentino & Sheppard (1978)	Swimming						
High affiliation motive		−1.11	24/20.00/4	2/2/1/1	1/1/1/1	3/1/2/1/1	3/6.00/1.00/3/2
Moderate affiliation motive		−0.46	26/20.00/4	2/3/1/1	1/1/1/1	3/2/2/1/1	3/6.00/1.00/3/2
Low affiliation motive		1.09	26/20.00/4	2/1/1/1	1/1/1/1	3/3/2/1/1	3/6.00/1.00/3/2
Stevenson (1990)	Brainstorming						
Experiment 1		1.05	96/20.00/4	1/1/2/1	1/2/1/1	2/2/3/2/2	2/8.00/8.00/3/2
Experiment 2							
High identifiability		0.10	96/20.00/4	1/2/2/1	1/2/1/1	3/2/3/2/2	2/8.00/8.00/3/2
Low identifiability							
Szymanski & Harkins (1987)	Brainstorming						
Experiment 1							
Self-evaluation		−0.52	34/20.00/4	1/2/2/1	1/2/1/1	3/2/3/2/2	2/—/1.00/3/2
No self-evaluation		1.06	34/20.00/4	1/1/2/1	1/2/1/1	2/2/3/2/2	2/—/1.00/3/2
Experiment 2							
Standard		−0.26	48/20.00/4	1/2/2/1	1/2/1/1	3/2/3/2/2	2/2.00/2.00/3/2
No standard		1.46	96/20.00/4	1/1/2/1	1/2/1/1	2/2/3/2/2	2/2.00/2.00/3/2
Thompson, Jackson, Williams, & Latané (1980)	Shouting						
High co-worker ability		0.22	40/20.00/4	1/1/1/2	1/1/1/1	2/2/3/2/1	3/4.00/4.00/3/2
Moderate co-worker ability		0.28	40/20.00/4	1/1/1/2	1/1/1/1	2/2/3/2/2	3/4.00/4.00/3/2
Low co-worker ability		0.13	40/20.00/4	1/3/1/2	1/1/1/1	2/2/3/2/3	3/4.00/4.00/3/2
Waller (1989)	Finding letters on sheet						
First and third graders		−0.19	150/ 7.78/1	1/1/2/1	1/3/1/1	4/2/4/2/2	3/6.00/6.00/3/2
Fifth graders		1.76	60/10.10/2	1/1/2/1	1/3/1/1	4/2/4/2/2	3/6.00/6.00/3/2
Weldon & Gargano (1985)	Job evaluation						
Experiment 1		1.54	23/20.00/4	1/1/2/2	3/4/2/2	2/2/3/2/2	2/16.00/1.00/1/2
Experiment 2		0.65	41/20.00/4	1/1/2/2	3/4/2/2	2/2/3/2/2	2/16.00/1.00/3/2
Weldon & Gargano (1988)	Job evaluation						
High accountability		0.10	26/20.00/4	1/2/2/2	3/4/2/2	3/1/3/2/2	2/16.00/17.50/3/2
Low accountability		1.41	24/20.00/4	1/1/2/2	3/4/2/2	2/2/3/2/2	2/16.00/17.50/3/2
Weldon & Mustari (1988), Experiment 1	Job evaluation						
Group size = 2		0.23	67/20.00/4	1/1/2/2	3/4/2/2	4/2/3/2/2	2/2.00/17.50/3/2
Group size = 16		1.17	67/20.00/4	1/1/2/2	3/4/2/2	4/2/3/2/2	2/16.00/17.50/3/2
Williams (1981)	Typing						
Cohesive groups		−0.51	15/21.50/4	2/2/1/1	1/1/1/1	2/1/1/2/2	3/4.00/1.00/2/2
Noncohesive groups		0.31	15/21.50/4	2/1/1/1	1/1/1/1	2/1/3/2/2	3/4.00/1.00/2/2
Williams & Burmont (1981)	Evaluating advertisements	0.73	30/20.00/4	1/1/2/1	1/2/1/1	2/2/3/2/2	2/3.00/3.00/3/2
Williams, Harkins, & Latané (1981)	Shouting						
Experiment 1							
High identifiability, group size = 2		0.12	48/20.00/4	1/2/1/2	1/1/1/1	3/2/3/2/2	3/2.00/6.00/1/2
High identifiability, group size = 6		0.55	48/20.00/4	1/2/1/2	1/1/1/1	3/2/3/2/2	3/6.00/6.00/1/2
Low identifiability, group size = 2		2.28	48/20.00/4	1/1/1/2	1/1/1/1	2/2/3/2/2	3/2.00/6.00/1/2
Low identifiability, group size = 6		2.72	48/20.00/4	1/1/1/2	1/1/1/1	2/2/3/2/2	3/6.00/6.00/1/2
Experiment 2							
Identifiable alone, group size = 2		1.40	36/20.00/4	1/1/1/2	1/1/1/1	2/2/3/2/2	3/2.00/4.00/1/2
Identifiable alone, group size = 4		1.89	36/20.00/4	1/1/1/2	1/1/1/1	2/2/3/2/2	3/4.00/4.00/1/2
Always identifiable, group size = 2		0.02	36/20.00/4	1/2/1/2	1/1/1/1	3/2/3/2/2	3/2.00/4.00/1/2
Always identifiable, group size = 4		0.05	36/20.00/4	1/2/1/2	1/1/1/1	3/2/3/2/2	3/4.00/4.00/1/2
Never identifiable, group size = 2		0.01	36/20.00/4	1/2/1/2	1/1/1/1	1/2/3/2/2	3/2.00/4.00/1/2
Never identifiable, group size = 4		0.19	36/20.00/4	1/2/1/2	1/1/1/1	1/2/3/2/2	3/4.00/4.00/1/2

Table 1 (*continued*)

Study and subset	Task	Effect size (*d*)	Subjects[a]	Methodology[b]	Task[c]	Theory-relevant predictors[d]	
Williams & Karau (1991)	Brainstorming						
Experiment 1							
High trust		1.29	43/20.00/4	1/1/2/1	1/2/1/1	2/2/3/2/1	2/6.00/6.00/3/2
Moderate trust		0.61	41/20.00/4	1/1/2/1	1/2/1/1	2/2/3/2/2	2/6.00/6.00/3/2
Low trust		−0.59	42/20.00/4	1/2/2/1	1/2/1/1	2/2/3/2/3	2/6.00/6.00/3/2
Experiment 2							
High co-worker effort		1.04	19/20.00/4	1/1/2/1	1/2/1/1	2/1/3/2/1	2/2.00/2.00/3/2
Low co-worker effort		−0.51	20/20.00/4	1/2/2/1	1/2/1/1	2/1/3/2/3	2/2.00/2.00/3/2
Experiment 3							
High meaningfulness, high co-worker ability		0.57	20/20.00/4	1/1/2/1	1/2/1/1	2/1/3/2/1	2/2.00/2.00/3/2
High meaningfulness, low co-worker ability		−1.15	20/20.00/4	1/2/2/1	1/2/1/1	2/1/3/2/3	2/2.00/2.00/3/2
Low meaningfulness, high co-worker ability		0.79	20/20.00/4	1/1/2/1	1/2/1/1	2/3/3/2/1	2/2.00/2.00/3/2
Low meaingfulness, low co-worker ability		1.29	20/20.00/4	1/1/2/1	1/2/1/1	2/3/3/2/3	2/2.00/2.00/3/2
Williams, Nida, Baca, & Latané (1989)	Swimming						
High identifiability		−0.77	8/20.00/4	2/2/1/1	1/1/1/1	3/1/1/1/1	3/4.00/4.00/1/2
Low identifiability		0.36	8/20.00/4	2/1/1/1	1/1/1/1	4/1/1/2/1	3/4.00/4.00/1/2
Williams & Williams (1981)	Shouting and clapping						
Experiment 1							
Kindergartners		0.11	112/5.00/1	2/3/1/1	1/1/1/1	2/2/4/2/2	3/2.00/2.00/3/2
Fourth and sixth graders		0.48	160/10.00/2	2/1/1/1	1/1/1/1	2/2/4/2/2	3/2.00/2.00/3/2
Experiment 2	Drawing circles	−0.53	20/ 3.50/1	2/2/1/1	1/1/1/1	2/2/2/2/2	3/20.00/20.00/3/2
Williams, Williams, Kawana, & Latané (1984)	Shouting and clapping						
Experiment 2		0.60	48/12.50/3	1/3/1/1	1/1/1/1	2/2/4/2/2	3/2.00/2.00/3/1
Experiment 3		0.87	16/20.00/4	1/3/1/1	1/1/1/1	2/2/3/2/2	3/2.00/2.00/3/1
Experiment 4		0.30	20/45.00/5	1/3/1/1	1/1/1/1	2/2/4/2/2	3/2.00/2.00/1/1
Yamaguchi, Okamoto, & Oka (1985)	Button pushing						
Experiment 1		*ns*	26/20.00/4	1/1/2/1	1/1/1/1	2/2/3/2/2	3/2.00/2.00/1/1
Experiment 2		*ns*	26/20.00/4	1/1/2/1	1/1/1/1	2/2/3/2/2	3/2.00/2.00/2/1

Note. Positive effect sizes indicate that subjects worked harder coactively than collectively (i.e., engaged in social loafing).

[a] The first variable is number of subjects; the second variable is mean age of subjects; the third variable is status of subjects (1 = third graders and younger, 2 = fourth through sixth graders, 3 = junior high or high school students, 4 = college students, 5 = adult nonstudents).

[b] The first variable is setting (1 = laboratory, 2 = field); the second variable is whether social loafing was predicted (1 = yes, 2 = no, 3 = no prediction or unclear); the third variable is design (1 = within-subjects, 2 = between-subjects); the fourth variable is cover story (1 = effort or performance related, 2 = other).

[c] The first variable is task complexity (1 = simple or well-learned, 2 = complex or novel, 3 = unclear); the second variable is type of effort required by task (1 = physical, 2 = cognitive, 3 = perceptual, 4 = evaluative); the third variable is quantity/quality emphasis of task (1 = maximizing, 2 = optimizing, 3 = mixed or unclear); the fourth variable is method of combining individual inputs (1 = additive, 2 = compensatory, 3 = other, mixed, or unclear).

[d] For the first set, the first variable is evaluation potential (1 = none, 2 = coactive only, 3 = coactive and collective, 4 = unknown or unclear); the second variable is task valence (1 = high, 2 = unspecified, 3 = low); the third variable is group valence (1 = high, 2 = moderate, 3 = low, 4 = unknown or unclear); the fourth variable is presence of group-level comparison standard (1 = yes, 2 = no); the fifth variable is expectations of co-worker performance (1 = high, 2 = unspecified, 3 = low). For the second set, the first variable is uniqueness of individual inputs to collective product (1 = unique, 2 = potentially redundant, 3 = completely redundant); the second variable is mean group size; the third variable is mean number of task performers present at session; the fourth variable is sex of subjects (1 = men, 2 = women, 3 = both); the fifth variable is culture (1 = Eastern, 2 = Western).

ANOVA) and a test of the homogeneity of the effect sizes within each class, Q_{Wi}. The continuous models, which are least squares linear regressions calculated with each effect size weighted by the reciprocal of its variance, provide a test of the significance of the predictor as well as a test of model specification (Hedges & Olkin, 1985). Finally, as a supplementary procedure, we conducted outlier analyses to determine whether the effect sizes were relatively homogeneous aside from the presence of a limited number of aberrant values (see Hedges, 1987; Hedges & Olkin, 1985).

Results and Discussion

Characteristics of Studies

Before considering the findings of the social loafing literature, we first present the characteristics of the studies from which our conclusions will be drawn. As shown by the central tendencies of the characteristics presented in Table 2, studies generally (a) were published recently, (b) were published as journal articles, (c) used a moderate number of subjects, (d) used college-age samples, and (e) only occasionally included organizational or nonadult samples. In addition, studies (a) were generally laboratory experiments with some field studies, (b) used both within- and between-subjects comparisons of individuals' effort across coactive and collective conditions, and (c) typically informed subjects that the purpose of the experiment was to examine effort or task performance. Regarding task characteristics, studies (a) typically used simple tasks, (b) tended to measure either physical or cognitive effort with some studies of perceptual or evaluative effort, and (c) generally used maximizing, additive tasks.

Finally, regarding key theory-relevant predictors, the last 10 variables in Table 1 show that studies generally (a) varied evalua-

Table 2
Summary of Study Characteristics

Variable and class	Value	Variable and class	Value
Median date of publication	1985	Conditions under which individual outputs can be evaluated	
Publication form		None	5
Journal article	121	Coactive condition only	117
Meeting paper	23	Coactive and collective conditions	28
Dissertation or master's thesis	14	Unclear	16
Unpublished document	8	Task valence	
Median number of subjects	48	High	30
Median age of subjects (years)	20	Moderate or no basis for determining	131
Status of subjects		Low	5
Third grade students or younger	6	Group valence	
Fourth through sixth grade students	5	High	9
Junior high or high school students	6	Moderate	11
College students	143	Low	129
Organizational employees	6	Unknown or unclear	17
Setting of study		Opportunity for group evaluation	
Laboratory	142	Present	12
Field	24	Absent	151
Type of coactive–collective comparison		Expectations of co-worker performance	
Within-subjects	69	High	14
Between-subjects	97	Moderate or no basis for determining	144
Type of cover story used		Low	8
Effort or performance related	115	Uniqueness of individual contributions to collective outcome	
Other	51	Unique	8
Complexity of task		Potentially redundant	64
Simple	135	Completely redundant	94
Complex	7	Median group size	3
Unknown or unclear	24	Median number of task performers present at session	3
Type of effort measured		Sex of subjects[a]	
Physical	70	Men	28
Cognitive	53	Women	12
Perceptual	29	Both	126
Evaluative	14	Culture of subjects	
Quantity or quality emphasis of task		Eastern	17
Maximizing	127	Western	148
Optimizing	30		
Mixed or unclear	9		
Method of combining individual inputs			
Additive	147		
Compensatory	11		
Other, mixed, or unclear	8		

Note. For categorical variables, numbers in table represent frequency of effort comparisons in each class. Summaries of continuous variables are based on reports for which information was available on each variable.

[a] When studies were subdivided on sex of subjects, the frequencies were 40 for men, 24 for women, and 114 for both.

tion potential across the coactive and collective conditions, (b) did not manipulate task valence, (c) studied low-valence groups of strangers, (d) did not provide an opportunity for group evaluation, (e) did not manipulate expectations of co-worker performance, (f) used tasks in which individuals' inputs to the collective product were either potentially redundant or completely redundant with the inputs of other group members, (g) had a relatively small median group size, (h) tended to have all task performers physically present at the experimental session, (i) typically included both male and female subjects, and (j) generally examined subjects from Western cultures.

Overall Magnitude of Social Loafing

The summary provided in Table 3 allows one to determine whether, on the whole, individuals tended to engage in social loafing and exert less effort collectively than coactively. An overall difference in effort across coactive and collective conditions is shown by a mean effect size that differed significantly from the 0.00 value that indicates exactly no difference (i.e., by a confidence interval that does not include 0.00). Consistent with the CEM and with prior reviewers' conclusions, individuals tended to engage in social loafing, producing lower effort levels in the collective condition than in the coactive condition. The weighted mean was very similar to the unweighted mean, whereas the median was somewhat lower. As shown by the homogeneity statistic given in Table 3, the effect sizes were not homogeneous (i.e., consistent) across the studies. A very large proportion (39%) of the effect sizes had to be removed to attain homogeneity, and the removal of outliers substantially decreased the value of the weighted mean. Inspection of the outliers revealed that among the 64 effect sizes removed, 49 were above and only 15 were below the median d computed for the original sample. The confidence interval for the weighted mean after outlier removal showed that the overall tendency toward decreased collective effort was still significant after outliers were removed.

Table 3
Summary of Effect Sizes (ds)

Criterion	Value
ds with outliers	
Sample size (n)	163
Mean weighted d (d_+)[a]	0.44
95% CI for d_+	0.39/0.48
Homogeneity (Q) of *ds* making up d_+[b]	964.70**
Mean unweighted d	0.48
95% CI for mean unweighted d	0.36/0.61
Median d	0.41
Differences indicating social loafing	128/163 (.79)
ds excluding outliers	
Sample size (n)	99
n removed outliers[c]	64 (.39)
Mean weighted d (d_+)	0.24
95% CI for d_+	0.19/0.29
Homogeneity (Q) of *ds* making up d_+	125.79

Note. Positive effect sizes indicate that subjects worked harder coactively than collectively (i.e., engaged in social loafing). CI = confidence interval; d = effect size; d_+ = mean weighted effect size; Q = homogeneity of *ds*.
[a] Effect sizes were weighted by the reciprocal of the variance. [b] Significance indicates rejection of the hypothesis of homogeneity. [c] The proportion appears in parentheses.
** $p < .001$.

The implication of the finding that people tend to engage in social loafing when working collectively depends in part on its magnitude. The several measures of central tendency reported in Table 3 are relevant to this issue. Taken together, these figures indicate approximately a four-tenths standard deviation in the direction of reduced collective effort. One means for interpreting the magnitude of the mean social loafing effect size is to compare it with effect sizes produced by other quantitative reviews in similar domains. Eagly's (1987) overview of a variety of mean effect sizes found in meta-analyses of social psychological research topics suggested that they ranged from about 0.00 to 1.20. In a particularly relevant comparison, Bond and Titus's (1983) review of the social facilitation literature reported that effect sizes ranged between −0.23 and 0.21, depending on the complexity of the task and the type of performance measure. Another standard is provided by Cohen (1977), who suggested that *d*s of 0.20 be labeled as *small*, *d*s of 0.50 be labeled as *moderate*, and *d*s of 0.80 be labeled as *large*. On the basis of these standards, social loafing might be regarded as small to moderate in magnitude, within the middle range of effects in the domain of social behavior, and larger than social facilitation.

However, it should be recognized that the tendency of researchers to focus on isolating conditions under which social loafing can be reduced or eliminated is likely to lead to an underestimation of the magnitude of the effect. Most studies were designed to test the potential of one or more factors for eliminating social loafing. Therefore, many of the coactive–collective comparisons included in our meta-analysis were drawn from situations in which social loafing was not predicted. Thus, the mean effect size of $d = 0.44$ may not necessarily provide the best measure of the overall magnitude of social loafing. Indeed, a categorical model computed for whether social loafing was predicted by the original researcher(s) was highly significant, $Q_B(2) = 310.86$, $p < .001$. Effect sizes were substantially larger ($d_+ = 0.76$, 95% CI = 0.70 to 0.81, $n = 89$) when social loafing was predicted than when social loafing was not predicted ($d_+ = -0.01$, 95% CI = −0.08 to 0.06, $n = 59$, contrast $p < .001$) and when researchers did not specify a prediction ($d_+ = 0.19$, 95% CI = 0.09 to 0.29, $n = 15$, post-hoc contrast $p < .005$).[5] The "predicted" and "not predicted" values may represent estimates of upper and lower bounds for the magnitude of social loafing. In view of these comparisons, the available research seems to suggest that there is at least a moderate tendency for individuals to engage in social loafing and to reduce their effort when working on collective tasks.

Finally, with regard to the robustness of social loafing, an

[5] Contrasts were a priori, unless otherwise indicated. The post-hoc contrast procedure is modeled after Scheffé's method for multiple comparisons in an analysis of variance (see Hedges & Olkin, 1985).

examination of the consistency of findings across studies is worthwhile. The direction of the effect was very consistent, evidenced by 79% of the comparisons being in the direction of reduced collective effort. However, outlier analyses showed that the effect sizes were highly inconsistent in magnitude: 39% of these effect sizes had to be removed to attain homogeneity. This figure is substantially larger than that reported in other meta-analyses of social behavior. It is possible that the focus of social loafing research on limiting conditions, combined with the wide range of tasks and subject populations used, contributed to this relative inconsistency. It is also possible that this inconsistency reflects the complex nature of the phenomenon and suggests that social loafing may represent a class of effects with psychologically distinct causal mechanisms.

Theory-Relevant Predictors of Social Loafing

The study attributes were examined as predictors of differences in individuals' coactive versus collective effort. Categorical models that yielded significant between-classes effects are presented in Table 4, and significant continuous models are presented in Table 5.

Evaluation potential. Consistent with the CEM and with evaluation accounts, the tendency for individuals to engage in social loafing decreased when evaluation potential was held constant across the coactive and collective conditions. Thus, individuals were more likely to loaf when their outputs could only be evaluated in the coactive condition than when their outputs could not be evaluated at all ($p < .001$) or when their outputs could be evaluated in both the coactive and collective conditions ($p < .001$). In fact, social loafing was eliminated when evaluation potential was not varied across coactive and collective conditions, as evidenced by the confidence intervals presented in Table 4. These findings are consistent with the claim made by Harkins (1987) and others that evaluation potential is a mediator of social loafing.

Task valence. As hypothesized, the tendency to engage in social loafing decreased as task valence increased. Contrasts showed that the tendency to loaf was greater when task valence was low than when task valence was either high ($p < .001$) or unspecified ($p < .06$, post-hoc contrast). Also, the tendency to loaf was greater when task valence was unspecified than when task valence was high ($p < .001$, post-hoc contrast). An examination of confidence intervals shows that individuals loafed when task valence was low or unspecified, but not when it was high. This suggests that high levels of task meaningfulness or personal involvement might eliminate loafing. Also, given that the unspecified tasks were likely relatively moderate in valence, these results suggest a direct relationship between task valence and individual effort on collective tasks.

Group valence and group-level comparison standards. Individuals did not loaf when group valence was high, but loafed in all other conditions. Thus, as predicted by the CEM and prior theories of self-validation processes in groups, the tendency to loaf was lower when group valence was high than when it was moderate ($p < .01$), unknown or unclear ($p < .05$, post-hoc contrast), or low ($p < .001$). Individuals also loafed less when group valence was either moderate ($p < .001$) or unknown ($p < .005$, post-hoc contrast) than when it was low. Finally, individuals loafed less when group valence was high than when it was low ($p < .001$). Similarly, the tendency to engage in social loafing was reduced when participants were provided with a group-level comparison standard. Taken together, these findings suggest that enhancing group cohesiveness or group identity might reduce or eliminate social loafing and that group-level outcomes are indeed valued by individuals in certain situations.

Expectations of co-worker performance. Consistent with the predictions of the CEM, individuals loafed when they expected their co-workers to perform well or when no expectations were provided, but did not loaf when they expected their co-workers to perform poorly (low vs. high contrast, $p < .005$; low vs. unspecified contrast, $p < .001$, post hoc). These findings, however, do not completely resolve the controversy between the matching of effort (Jackson & Harkins, 1985) and social compensation (Williams & Karau, 1991) views. The basic pattern of findings is consistent with both the CEM and the social compensation research, which both suggest that individuals' outcomes are more contingent on their efforts when working with others who are expected to perform poorly. However, effort levels were not significantly higher collectively than coactively when participants expected their co-workers to perform poorly, as has been found in the prior social compensation work.

Nevertheless, a resolution is possible when the role of task meaningfulness is considered. Because social compensation represents a motivation gain, individuals must not only increase their usual collective effort, they must actually work harder than they would ordinarily work alone. The logic of the CEM suggests that individuals would be unlikely to exert such extraordinary effort unless they viewed the task as meaningful and therefore highly valued a favorable outcome. In support of this notion, when the interaction between expectations of co-worker performance and task meaningfulness was considered, a significant motivation gain was found only for the four comparisons in which individuals worked on a meaningful task with co-workers who were expected to perform poorly ($d_{i+} = -0.69$, 95% CI = -1.11 to -0.27). These results highlight the potential of the CEM for predicting situations in which motivation gains might emerge on collective tasks.

Uniqueness of individual inputs. Individuals worked just as hard collectively as coactively when their individual inputs to the collective product were unique, but loafed when their inputs were either potentially redundant or completely redundant ($ps < .001$). Consistent with the CEM and with the results of Kerr (1983; Kerr & Bruun, 1983) and Harkins and Petty (1982), individuals may perceive that their efforts are less instrumental in obtaining valued outcomes when their inputs to the collective product are either dispensable or largely overlapping with those of others.[6]

[6] Although uniqueness and dispensability are often closely related, they are conceptually distinct. A group member's inputs can be indistinguishable from the inputs of others, yet essential to group performance. For example, on an additive rope-pulling task, group members'

Group size. Simple linear regressions showed that both group size and the number of task performers present were positively related to social loafing: Effect sizes were larger for studies that combined more individual inputs into the collective product and for studies that had more individuals perform the task at each session. Most of the theories discussed in the introduction would predict group size effects, albeit for different reasons. Social impact theory and self-attention theory devote particular attention to group size effects relative to other aspects of social loafing.

Gender and culture. Consistent with the CEM, the magnitude of social loafing was larger in studies that used samples of only male subjects than for studies that used either mixed samples ($p < .05$, post-hoc contrast) or samples of only female subjects ($p < .001$). If this difference were indeed attributable to a greater tendency by men to focus on individualistic outcomes, the CEM would predict that men would be more attentive to strategic concerns than women when working on collective tasks. Consistent with this reasoning, the group size effect reported earlier was only reliable for comparisons involving groups of men ($b = .13$, $b^* = .36$, $p < .001$). This finding suggests that men may be particularly attentive to the reduced efficacy of their inputs when working in larger groups, whereas women may tend to exert more consistent levels of collective effort across various group sizes.

Also as predicted, the magnitude of social loafing was larger for subjects from Western cultures than for subjects from Eastern cultures. These provocative findings suggest that both women and individuals in Eastern cultures are less likely to engage in social loafing, presumably because they have more group-oriented priorities than do men and individuals in Western cultures.

Task complexity. Individuals performed better coactively

magnitude of social loafing was also larger in laboratory experiments than in field studies. Interestingly, the magnitude of the effect did *not* vary depending on whether (a) the coactive–collective comparison was made within-subjects or between-subjects. This finding suggests that social loafing does not occur simply because subjects allocate greater effort to coactive trials than to collective trials (see Harkins, Latané, & Williams, 1980; Kerr & Bruun, 1981). Finally, the magnitude of social loafing was not dependent on whether (a) maximizing or optimizing tasks were used, (b) tasks were additive or compensatory, and (c) physical, cognitive, evaluative, or perceptual effort was measured, suggesting that social loafing is robust across tasks.

Impact of Other Variables on Social Loafing

The trend toward decreased collective effort was enhanced when a deceptive cover story was used. Thus, subjects were more likely to loaf when they were not informed that the purpose of the study was to examine effort or performance, suggesting that telling subjects the purpose of the experiment may inhibit their usual tendency to engage in social loafing. Social loafing was also greater for studies reported in journal articles than for dissertations and master's theses or for unpublished documents ($ps < .05$, post-hoc contrasts). Finally, as shown in Table 5, a simple linear regression showed that date of publication was negatively related to the magnitude of social loafing such that effect sizes were larger for earlier studies. However, interpretation of this time trend is clouded when it is recognized that the earlier studies both (a) tended to use fewer experimental paradigms that were more easily controlled, possibly resulting in smaller error terms and larger effect sizes and (b) tended to focus on documenting the effect rather than isolating moderating variables and were thus more likely to find larger effect sizes across all comparisons.