

# When Small Effects Are Impressive

Deborah A. Prentice and Dale T. Miller  
Princeton University

Effect size is becoming an increasingly popular measure of the importance of an effect, both in individual studies and in meta-analyses. However, a large effect size is not the only way to demonstrate that an effect is important. This article describes 2 alternative methodological strategies, in which importance is a function of how minimal a manipulation of the independent variable or how difficult-to-influence a dependent variable will still produce an effect. These methodologies demonstrate the importance of an independent variable or psychological process, even though they often yield effects that are small in statistical terms.

Psychologists are increasingly interested in statistical techniques that allow them to say something about the importance of their effects. This growing interest stems in large part from the realization that conventional significance-testing procedures provide an impoverished and possibly even misleading view of how seriously to take any particular result. Current wisdom regarding the use of statistics in psychological research holds that (a) the size of an effect is at least as informative as its statistical significance, if not more informative and (b) meta-analysis provides an important tool for assessing the reliability and magnitude of an effect across multiple studies (see, e.g., Cohen, 1990; Rosnow & Rosenthal, 1989). Underlying these points is the general argument that one should pay attention to size, as well as significance level, in deciding how impressed to be with an effect.

Whereas the use of effect size and other statistical measures of strength is relatively new in psychology, the goal of demonstrating the importance of an effect is not new at all. In this article, we examine the alternative ways in which psychologists have approached this task and the implications of these approaches for questions of how much variance is accounted for. We argue that what makes some effects seem important is not their magnitude but rather the methodologies of the studies that produced them. The statistical size of an effect is heavily dependent on the operationalization of the independent variables and the choice of a dependent variable in a particular study. Thus, with sufficient ingenuity, a researcher can design an experiment so that even a small effect is impressive.

Our purpose here is to document these methodological strategies for demonstrating important effects. We consider effects to be important to the extent that they have had a major impact on thinking in the field (e.g., findings that are frequently cited, those that are featured in survey textbooks). Thus, our analysis is retrospective; we focus on examples of studies that have pro-

vided convincing demonstrations of the importance of certain psychological variables or processes, despite the fact that many of them have yielded small effects. Moreover, we make no assumptions about the motivations or intentions of the researchers whose work we cite but simply seek to make explicit the methodological approaches that they have used so successfully. We begin with a brief review of the rationale for using measures of effect size as an index of importance and then describe two alternative methodological strategies for demonstrating an important effect.

## Statistical Strength of an Effect

One reasonable way to determine the importance of an effect is to compute it, using one of a family of effect-size measures (Cohen, 1977). The two most commonly used measures of effect size are the standardized mean difference ( $d$ ) and the correlation coefficient ( $r$ ), although there is an effect size index appropriate to any statistical test. These measures have many beneficial properties: (a) They indicate the degree to which a phenomenon is present in a population on a continuous scale, with zero always indicating that the phenomenon is absent (i.e., that the null hypothesis is true), (b) they come with conventions for what values constitute a small, medium, and large effect, (c) they provide some indication of the practical significance of an effect (which significance tests do not), (d) they can be used to compare quantitatively the results of two or more studies, and (e) they can be used in power analyses to guide decisions about how many subjects are needed in a study (see Cohen, 1977, 1990; Rosnow & Rosenthal, 1989). In short, effect size is a simple, easy-to-understand quantitative measure that provides one useful index of the importance of an effect.

An additional argument in favor of using effect size as a measure of importance is that effect sizes can be collected across studies. Most contemporary approaches to meta-analysis involve estimating effect sizes for each of a set of relevant studies or findings and then analyzing the mean and variability of these estimates (see Bangert-Drowns, 1986). Thus, effect size can serve as a measure of the importance of an effect not only in the context of a single study but also in a review of multiple studies conducted within a similar paradigm. For this reason,

---

We wish to thank Robert Abelson, Carl Sherrick, and participants in the social research seminar at Princeton University for comments on drafts of this article.

Correspondence concerning this article should be addressed to Deborah A. Prentice or Dale T. Miller, Department of Psychology, Princeton University, Green Hall, Princeton, New Jersey 08544-1010.

many researchers have suggested that effect sizes should be reported routinely for all significant and nonsignificant results (see Rosnow & Rosenthal, 1989).

### Alternative Methods of Demonstrating the Importance of an Effect

Effect size and other measures of variance accounted for are unquestionably useful for assessing the magnitude of an effect and serve as an important supplement to conventional significance tests. One might question, in fact, why it has taken psychologists so long to discover these procedures (Cohen, 1990). One possible answer to this question is that in some areas of psychology, researchers have relied on alternative conceptions of what makes an effect seem important. Whether intentionally or unintentionally, these researchers have approached the problem of how to demonstrate the importance of an effect with more attention to design than to analysis: They have adopted methodological strategies that create impressive demonstrations, even though the studies often yield effects that are statistically small. We consider two of these strategies, along with their implications for statistical measures of strength.

#### *Minimal Manipulations of the Independent Variable*

One strategy for demonstrating important effects involves showing that even the most minimal manipulation of the independent variable still accounts for some variance in the dependent variable. A classic example of this approach is the so-called minimal group experiments of Tajfel and his colleagues (e.g., Billig & Tajfel, 1973; Tajfel, Billig, Bundy, & Flament, 1971). At the time these experiments were conducted, much research had already demonstrated that people favor members of their own group over members of other groups. But these investigators were interested in identifying the minimal conditions necessary to produce this ethnocentrism effect and thus conducted a series of studies using increasingly minimal manipulations of group membership. In one of the early studies in this series, boys were told that they tended either to overestimate or to underestimate the number of dots on briefly presented slides (Tajfel et al., 1971). When later given the opportunity to allocate points in a game, overestimators consistently allocated more points to other overestimators and underestimators to other underestimators. This effect was taken as strong evidence of ethnocentrism: Even though the groups were based on a meaningless classification and members had no contact with each other, they still showed a preference for the in group.

Subsequent minimal group experiments provided still more convincing evidence of the importance of ethnocentrism without yielding effects of any greater magnitude. In the most minimal of the experiments, subjects were told that they were being assigned to groups at random and were even shown the lottery ticket that determined whether they were a member of the Phi group or the Gamma group (Locksley, Ortiz, & Hepburn, 1980). Even with explicit random assignment, subjects still showed a preference for members of their own group. The minimal group experiments, and this last study in particular, are impressive demonstrations of ethnocentrism, regardless of the size of the effects they produce. Indeed, the strength of these

demonstrations derives not from the proportion of variance in allocations that group membership can account for but instead from the fact that such a slight manipulation of group membership can account for any variance in allocations at all.

Another example of this methodological tradition is provided by research on the effects of mere exposure on liking (see Harrison, 1977, for a review). Studies have demonstrated that exposure increases liking for stimuli as diverse as musical selections, Chinese-like characters, photographs of men's faces, and nonsense words, both in laboratory (Zajonc, 1968) and field (Zajonc & Rajecki, 1969) experiments. But just how mere an exposure is necessary to show increased liking? Additional investigations have focused on exploring the limits of this mere exposure effect. In one study, subjects listened to an audiotape of a prose passage in one ear while musical melodies played in their other, unattended ear. Even though they could not recognize the melodies later, subjects still liked them better than melodies to which they had not been exposed (Wilson, 1979). In another study, subjects were shown slides of geometric figures for durations too brief to permit recognition and still preferred these figures to those they had not previously seen (Kunst-Wilson & Zajonc, 1980). The minimal manipulations used in these studies did more than just provide yet another demonstration of the exposure-liking effect; they also showed how simply and subtly this effect could be produced.

The psychological literature (particularly the social psychological literature) offers many more examples of the minimalist approach to demonstrating an important effect. In these studies, the use of a minimal manipulation serves to demonstrate that even under the most inauspicious circumstances, the independent variable still has an effect. Consider, for example, a study by Isen and Levin (1972) that showed that putting people in a good mood leads them to be more helpful. They manipulated mood by giving some subjects cookies while they studied in the library (good mood) and giving other subjects nothing (control). There are clearly many stronger manipulations of mood that they might have used. They could, perhaps, have given good-mood subjects a free meal in a fancy restaurant or good grades in their courses or even a winning ticket in the lottery. These manipulations may very well have shown a stronger effect of mood on helping in terms of variance accounted for.<sup>1</sup> But Isen and Levin's cookie study still provides a convincing and memorable demonstration of the effect; the power of this demonstration derives in large part from the subtlety of the instigating stimulus. Indeed, this demonstration would become no less impressive if a meta-analysis on cookie studies showed that the manipulation accounted for little variance. Furthermore, although mood effects might be interesting however heavy-handed the manipulation that produced them, the cookie study was perhaps made more interesting by its reliance on the minimalist approach.

<sup>1</sup> We do not imply here that when a small manipulation produces a small effect, a large manipulation will always produce a large effect. Indeed, a linear relationship between the size of the manipulation and the size of the effect is not necessary to our claims about the importance of small effects. We argue that a result can be important regardless of its magnitude if it changes the way people think about a psychological variable or process.

### *Choice of a Difficult-to-Influence Dependent Variable*

A second approach to demonstrating important effects involves choosing a dependent variable that seems especially unlikely to yield to influence from the independent variable. A good example of this strategy comes from the literature on physical attractiveness. Many studies have shown that physically attractive people are seen as more intelligent, successful, sociable, kind, sensitive, and so on (see Berscheid & Walster, 1974 for a review). These findings suggest that physical attractiveness has a powerful effect on social perception. Even more convincing evidence of the importance of this effect comes from studies showing that physically attractive people receive more positive job recommendations, even when attractiveness could not possibly influence job performance (Cash, Gillen, & Burns, 1977). But could we imagine a still more impressive demonstration of the importance of physical attractiveness in social perception? Efran (1974) examined the effect of the physical attractiveness of a defendant on judgments of guilt and severity of punishment by a simulated jury. Even though legal judgments are supposed to be unaffected by such extraneous factors as attractiveness, in fact, Efran found that attractive defendants were judged less likely to be guilty and received less punishment than unattractive defendants (see also Sigall & Ostrove, 1975). This demonstration that physical attractiveness matters in the courtroom is impressive, despite the fact that it matters much less here than in other domains of interpersonal judgment. One is inclined to conclude from this study that if attractiveness can even affect legal judgments, then there is no domain of social perception that is immune to its influence.<sup>2</sup>

Another example of achieving a convincing demonstration through selection of a resistant dependent variable is Asch's (1951) classic studies of conformity to group pressure. At the time that Asch undertook these studies, much research had already demonstrated the influence of group pressure on perceptual judgments when reality was ambiguous (e.g., Sherif, 1936). Asch believed that a truer test of the power of group pressure would require individuals to yield to a group judgment that they "perceived to be contrary to fact" (Asch, 1951, p. 177). In a prototypical study, a naive subject was asked to judge the length of a line after observing each of 8 other subjects (who were actually experimental confederates) give the same objectively incorrect answer. In this situation, one third of the judgments of naive subjects conformed with the erroneous judgment of the majority. This finding provides a striking demonstration of the importance of group pressure, regardless of whether one considers one third a large effect or a small effect. The fact that any subjects conformed to an obviously incorrect judgment is impressive.<sup>3</sup>

This strategy of showing that a psychological variable or process is important by demonstrating that it operates even in domains you would think were immune to its effects goes beyond the experimental tradition. For example, Durkheim's (1897/1951) finding of a relationship between social structure and suicide rates was impressive despite the fact that these macro variables surely cannot account for much of the variance. But the strength of the finding derives from the implication that if a behavior as individualistic and atomistic as suicide is correlated with social structure, we cannot assume that there is any micro-

behavior that is independent of it. Similarly, Freud's (1901/1971) analysis of the psychopathology of everyday life strongly suggested a pervasive influence of unconscious motives even though the incidence of slips of the tongue and lapses of memory is quite low. Again, the argument is that if the unconscious intrudes even in ordinary speech and memory, it must be quite powerful indeed.

Before leaving this section, we should note that judgments of the importance of an effect are, of course, highly subjective. Moreover, our arguments for the impressiveness of the demonstrations we have described apply primarily to researchers who focus on the independent variables or psychological processes under investigation, not to those who focus on the dependent variables. We would not, for example, expect a legal scholar to be impressed with the Efran study, nor would we necessarily expect a suicidologist to consider Durkheim's finding important. For investigators who define their research area in terms of a particular dependent variable or empirical relationship (i.e., convergent researchers; see McGuire, 1983), variance accounted for may very well be the critical measure of the importance of an effect.

### Statistical Versus Methodological Routes to an Impressive Demonstration

As we have suggested, statistical measures of variance accounted for are not the only tools researchers have to show that an effect is important. Despite the many virtues of these measures, in the context of particular studies, they can prove to be quite limited for conveying the importance of a finding. Declaring an effect to be important in effect-size terms is saying that a particular operationalization of the independent variable accounted for a lot of the variance in a particular dependent variable. This conception of importance makes sense if the experimenter is committed to the operations that were used to generate the data. If, however, the experimenter could easily have operationalized the independent variable differently or chosen a different dependent variable, the argument for using effect size, or more generally variance accounted for, as a measure of importance breaks down (see Mayo, 1978, for a similar argument).

In psychology, the utility of statistical versus methodological strategies for demonstrating the importance of an effect tends to divide along area lines. Statistical approaches are most useful in areas of psychology in which the operationalization of the independent variable and the choice of a dependent variable are clearly defined by the problem itself. For example, investigators interested in comparing the effectiveness of different

<sup>2</sup> Sudnow (1967) has suggested that physical attractiveness may even influence the speed with which people are pronounced dead on arrival in emergency rooms. An empirical demonstration of this effect might well be the most impressive evidence for the importance of physical attractiveness yet!

<sup>3</sup> The Asch experiments also demonstrated how minimal a manipulation of group pressure was required to produce the effect by using an ad hoc group composed entirely of individuals unknown to the subject and by showing that even with only 3 members of the majority group (compared with 8 in the prototypical case), the effect still held.

methods of classroom teaching, the outcomes of different psychotherapeutic techniques, or the validity of different aptitude tests are typically committed to their operationalizations of these variables and to their choice of outcome measures. In these cases, effect size is a perfectly appropriate measure of importance, and indeed, meta-analyses have proven very useful for reviewing studies in these areas (see Bangert-Drowns, 1986).

By contrast, the problems addressed in other areas of psychology afford the investigator a great deal more latitude in decisions regarding experimental design. Investigators of the effects of ethnocentrism, stimulus exposure, or mood have many possible operationalizations of these variables at their disposal. Similarly, those interested in demonstrating the importance of physical attractiveness or group pressure can choose among a multitude of dependent measures. Social psychologists who study these problems often design their studies so as to explore the limits of the effects. Studies in this tradition are likely to result in some number of small effect sizes and skeptical meta-analyses. But although these effect sizes may force us to reconsider the strength of an operationalization or the choice of a dependent variable (both of which were, in fact, designed to yield small effects), they do not force us to reconsider the importance of an independent variable or a psychological process.

One difficulty raised by these methodological approaches to demonstrating an important effect is how to quantify them.<sup>4</sup> That is, how does one measure just how minimal a manipulation is or how unlikely a dependent variable is to yield to influence? Although we know of no simple metric on which to rely, one possible strategy is to argue, using Bayesian reasoning, that an effect is important to the extent that it increases the odds that a hypothesis is true compared with its alternatives (Abelson, 1990). For example, consider the hypothesis that a good mood increases helping. The odds that this hypothesis is true might be enhanced to a greater extent by an experiment showing that cookie recipients help more than by an experiment showing that lottery winners help more. Similarly, the hypothesis that physical attractiveness affects social perception might become relatively more likely given a demonstration that attractiveness affects judgments in the courtroom than that it affects judgments in the personnel office. This strategy works well in principle, but unfortunately, the practical difficulties of applying Bayes's theorem (e.g., estimating prior probabilities; see Abelson, 1990) limit its utility. Still, this Bayesian approach highlights the fact that the amount of variance an effect accounts for is just one of many ways to think about its importance and further suggests the possibility that alternative conceptions of importance can be quantified.

We are not the first to argue that small effects can, in fact, be important. Three major defenses of their potential importance have been offered previously: (a) Small effects may have enormous implications in a practical context, (b) small effects in ongoing processes may accumulate over time to become large effects, and (c) small effects may be quite important theoretically (see, e.g., Abelson, 1985; Mook, 1983; Rosenthal & Rubin, 1983; Yeaton & Sechrest, 1981). These arguments are well-taken, but they differ in both spirit and substance from what we are asserting here. In the types of studies we have described, small effects are important not because they have practical con-

sequences nor because they accrue into large effects, nor because they lead to theory revision (indeed, in most of these cases, the effect or process under investigation was well established prior to the studies described). Instead, they are important because they show that an effect is so pervasive, it holds even under the most inauspicious circumstances.<sup>5</sup> Moreover, these methodological strategies for demonstrating importance underscore the fact that the size of an effect depends not just on the relationship between the independent and dependent variables but also on the operations used to generate the data. Many studies are not designed to account for a lot of variance and are no less impressive for the statistical size of the effects they produce.

## Summary and Conclusions

In summary, we have argued here that although effect size can be a very useful measure of the strength of an effect, there are alternative ways to demonstrate that an effect is important. We have focused on two methodological approaches, in which importance is a function of how minor a manipulation of the independent variable or how resistant a dependent variable will still produce an effect. Our purpose has been to make explicit what experimenters who have used these methodologies have perhaps known implicitly: Showing that an effect holds even under the most unlikely circumstances possible can be as impressive as (or, in some cases, perhaps even more impressive than) showing that it accounts for a great deal of variance. Indeed, researchers might do well to consider these alternative goals (e.g., accounting for maximal variance, using the most minimal manipulation) when designing and reporting their studies.

The arguments we have made against the exclusive use of effect size as a means for evaluating the importance of empirical results apply equally well to regression analysis, path analysis, and all other techniques that are based on calculation of the proportion of variance accounted for. These techniques can tell us a lot about the strength of a particular operationalization, but their utility as measures of importance is limited by the relation of that operationalization to the independent variable or psychological process under investigation. In the studies we have described, investigators have minimized the power of an

<sup>4</sup> In the examples in this article, we have set the criteria for a minimal manipulation of an independent variable and a difficult-to-influence dependent variable relative to other studies in the research area. For example, the minimal group studies use a minimal manipulation of group membership relative to other studies of ethnocentrism. However, one can conceive of these criteria more broadly in terms of people's expectations about whether a particular operationalization of an independent variable should have an effect or whether a particular dependent variable should be influenceable.

<sup>5</sup> Investigators have used a similar logic of showing that an effect holds even under inauspicious conditions by demonstrating an established effect on a population that seems very unlikely to be affected. For example, showing that even physicians are overconfident about their diagnoses (Christensen-Szalanski & Bushyhead, 1981) or that even divinity students will not stop to help an emergency victim (Darley & Batson, 1973) provides impressive evidence for these psychological phenomena.

operationalization and, in so doing, have succeeded in demonstrating the power of the underlying process. Thus, a small effect size, low multiple correlation, or negligible path value will not lead these investigators to question their conclusions. On the contrary, they will be pleased that their effect survived the toughest test they could give it and will be more convinced than ever of its importance.

### References

- Abelson, R. P. (1985). A variance explanation paradox: When a little is a lot. *Psychological Bulletin*, *97*, 128-132.
- Abelson, R. P. (1990). *Thinking about statistics*. Unpublished manuscript, Yale University.
- Asch, S. (1951). Effects of group pressure upon the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, leadership and men* (pp. 177-190). Pittsburgh, PA: Carnegie Press.
- Bangert-Drowns, R. L. (1986). Review of developments in meta-analytic method. *Psychological Bulletin*, *99*, 388-399.
- Berscheid, E., & Walster, E. (1974). Physical attractiveness. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 7, pp. 157-215). San Diego, CA: Academic Press.
- Billig, M., & Tajfel, H. (1973). Social categorization and similarity in intergroup behavior. *European Journal of Social Psychology*, *3*, 27-52.
- Cash, T., Gillen, B., & Burns, D. (1977). Sexism and "beautism" in personnel consultant decision making. *Journal of Applied Psychology*, *62*, 301-310.
- Christensen-Szalanski, J., & Bushyhead, J. (1981). Physicians' use of probabilistic information in a real clinical setting. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 928-935.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. San Diego, CA: Academic Press.
- Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, *45*, 1304-1312.
- Darley, J., & Batson, C. (1973). From Jerusalem to Jericho: A study of situational and dispositional variables in helping behavior. *Journal of Personality and Social Psychology*, *27*, 100-108.
- Durkheim, E. (1951). *Suicide* (J. Spaulding & G. Simpson, Trans.). Glencoe, IL: Free Press. (Original work published 1897)
- Efran, M. (1974). The effect of physical appearance on the judgment of guilt, interpersonal attractiveness, and severity of recommended punishment in a simulated jury task. *Journal of Research in Personality*, *8*, 45-54.
- Freud, S. (1971). *The psychopathology of everyday life* (A. Tyson, Trans.). New York: Norton. (Original work published 1901)
- Harrison, A. (1977). Mere exposure. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10, pp. 39-83). San Diego, CA: Academic Press.
- Isen, A. M., & Levin, P. F. (1972). The effect of feeling good on helping: Cookies and kindness. *Journal of Personality and Social Psychology*, *21*, 384-388.
- Kunst-Wilson, W., & Zajonc, R. (1980). Affective discrimination of stimuli that cannot be recognized. *Science*, *207*, 557-558.
- Locksley, A., Ortiz, V., & Hepburn, C. (1980). Social categorization and discriminatory behavior: Extinguishing the minimal intergroup discrimination effect. *Journal of Personality and Social Psychology*, *39*, 773-783.
- Mayo, R. J. (1978). Statistical considerations in analyzing the results of a collection of experiments. *Behavioral and Brain Sciences*, *1*, 400-401.
- McGuire, W. J. (1983). A contextualist theory of knowledge: Its implications for innovation and reform in psychological research. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 16, pp. 1-47). San Diego, CA: Academic Press.
- Mook, D. G. (1983). In defense of external invalidity. *American Psychologist*, *38*, 379-387.
- Rosenthal, R., & Rubin, D. (1983). A note on percent of variance explained as a measure of the importance of effects. *Journal of Applied Social Psychology*, *9*, 395-396.
- Rosnow, R., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist*, *44*, 1276-1284.
- Sherif, M. (1936). *The psychology of group norms*. New York: Harper & Row.
- Sigall, H., & Ostrove, N. (1975). Beautiful but dangerous: Effects of offender attractiveness and nature of the crime on juridic judgments. *Journal of Personality and Social Psychology*, *31*, 410-414.
- Sudnow, D. (1967). Dead on arrival. *Transaction*, *5*, 36-44.
- Tajfel, H., Billig, M., Bundy, R., & Flament, C. (1971). Social categorization and intergroup behavior. *European Journal of Social Psychology*, *1*, 149-178.
- Wilson, W. (1979). Feeling more than we can know: Exposure effects without learning. *Journal of Personality and Social Psychology*, *37*, 811-821.
- Yeaton, W., & Sechrest, L. (1981). Meaningful measures of effect. *Journal of Consulting and Clinical Psychology*, *49*, 766-767.
- Zajonc, R. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *Monograph Supplement*, *9*(2, Pt. 2).
- Zajonc, R., & Rajecki, D. (1969). Exposure and affect: A field experiment. *Psychonomic Science*, *17*, 216-217.

Received February 25, 1991

Revision received August 2, 1991

Accepted August 16, 1991 ■