A Meta-Analytic Critique of Relative Deprivation

Heather J. Smith
Thomas F. Pettigrew
Gina M. Pippin
Silvana Bialosiewicz

1 Sonoma State University of California
2 University of California, Santa Cruz

Corresponding Author:
Heather Smith, Department of Psychology, Sonoma State University
1801 East Cotati Ave., Rohnert Park, CA 94928

Email: smithh@sonoma.edu
Abstract

Relative deprivation (RD) is the judgment that one is worse off compared to some standard accompanied by feelings of anger and resentment. Social scientists use RD to predict a wide range of significant outcome variables: willingness to join protests, individual achievement and deviance, intergroup attitudes, and physical and mental health. But the results are often weak and inconsistent. To determine whether these results reflect measurement or theoretical deficiencies, the authors conducted a meta-analytic review of 210 studies (representing 293 independent samples and 186,073 respondents). RD measures that (1) include justice-related affect, (2) match the outcome level of analysis and (3) used higher quality measures yielded significantly stronger relationships. Future research should focus on appropriate RD measurement, affect and the inclusion of theoretically relevant appraisals of the situation. Such methodological improvements would revitalize RD as a useful social psychological predictor of a wide range of important individual and social processes.

Keywords: relative deprivation, meta-analysis, affect, protest participation, deviance, health, prejudice, social comparison, social justice
A Meta-Analytic Critique of Relative Deprivation

Marx (1847) captures the intuitive appeal of relative deprivation (RD) as an explanation for social behavior. If comparisons to other people, groups or even themselves at different points in time lead people to believe that they do not have what they deserve, they will be angry and resentful. Relative deprivation (RD) describes these subjective evaluations.

Thus, RD is a social psychological concept *par excellence*. It postulates a subjective state that shapes emotions, cognitions, and behavior. It links the individual with the interpersonal and intergroup levels of analysis. It melds easily with other social psychological processes to provide more integrative theory – a prime disciplinary need (Pettigrew, 1991). Moreover, RD challenges conventional wisdom about the importance of absolute deprivation for collective action, individual deviance and physical health. And it has proven useful in a wide range of areas. Researchers have invoked RD to explain phenomena ranging from poor physical health (Adler, Epel, Catellazzo & Ickovics, 2000) to participation in collective protest (Newton, Mann & Geary, 1980) and even susceptibility to terrorist recruitment (Moghaddam, 2005). Indeed, the concept has been used throughout the social sciences (Walker & Smith, 2001), from criminology (e.g., Lea & Young, 1993) and economics (e.g., Yitzhaki, 1979) to political science (e.g., Lichbach, 1990) and history (Snyder & Tilly, 1972).
Some investigations strongly support RD models (e.g., Abrams & Grant, 2010; Leach, Iyer & Pederson, 2007; Pettigrew, Christ, Wagner, Meertens, Van Dick & Zick, 2008; Pettigrew & Meertens, 1995; Vanneman & Pettigrew, 1972; Walker & Mann, 1987), but others do not (e.g., Gaskell & Smith, 1984; Macleod, Smith, Metcalfe & Hart, 2005; Schmitt, Maes, & Widaman, 2010; Snyder & Tilly, 1972; Thompson, 1989). In response to these inconsistencies, several previous literature reviews have sought to clarify the theoretical antecedents and components of the concept (Crosby, 1976; Martin, 1986a; Walker & Pettigrew, 1984). Other reviews, however, dismiss its value altogether (Brush, 1996; Finkel & Rule, 1987; Gurney & Tierney, 1982). We believe such dismissals of the concept are premature.

The purpose of this review is to determine whether the failure of RD to predict particular outcomes is the product of empirical or theoretical deficiencies. We propose that RD’s inconsistent results can largely be explained by measurement, rather than theoretical, problems. To test this contention across an array of outcome variables, this paper offers a quantitative literature review of the relevant social science research from 1949 to January 2010. Unlike traditional qualitative literature reviews, a meta-analytic integration of research results will enable us to determine whether the RD effects are as weak or nonexistent as some critics claim.

History of the Concept

Stouffer (1949) coined RD to describe unexpected relationships that emerged from surveys of American soldiers in World War II. For example, Stouffer and his colleagues found that U.S. Army Air corpsman reported more frustration over promotions in comparison to the military police even though they enjoyed a much faster rate of promotion than the police. Stouffer maintained that the military police were not the relevant comparison for the airmen; within their Air Corps ingroup, they knew many similar peers who had been promoted. The American Soldier researchers did not measure RD directly; rather, they inferred it as a post hoc explanation. This failure to initiate a
prototype measure has led to literally hundreds of diverse and sometimes conflicting measures that have bedeviled RD research ever since.

After Stouffer introduced RD, Merton (1957; Merton & Kitt, 1950) enlarged the idea within a reference group framework. This work led Pettigrew (1967) to point out that RD was but one of a large family of concepts and theories that employed relative comparisons in both sociological and psychological social psychology. From sociology, this theoretical family embraces Hyman's (1942, 1960) and Merton's (1957) reference group theory, Lenski's (1954) concept of status crystallization, Blau's (1964) concept of fair exchange, and Homans' (1961) concept of distributive justice. From psychology, these social evaluation ideas include Hatfield, Walster and Bersheid's (1978) equity theory, Festinger’s (1954) social comparison theory, and Thibaut and Kelley's (1959) concept of comparison level.

Runciman (1966) broadened the RD construct by distinguishing between egoistic (individual) and fraternal (group) RD. A person could believe that she is personally deprived (individual RD or IRD) or that a social group to which she belongs is deprived (group RD or GRD). Feelings of GRD should be associated with group serving attitudes and behavior such as collective action and outgroup prejudice, whereas IRD should be associated with individual-serving attitudes and behavior such as academic achievement and property crime.

During the following decades, scholars incorporated RD into larger models of social comparison, casual attribution and equity theory. In one model, Crosby (1982) proposed that IRD requires (1) wanting what one does not have, and (2) feeling that one deserves whatever it is one wants but does not have. In a second model, Folger (1987) proposed that a person’s current situation forms a narrative or story to which different alternative stories can be compared. People will evaluate their current outcome negatively and feel resentful and angry if they can imagine (1)
better alternative outcomes, (2) more legitimate procedures that could produce better outcomes, and (3) the current situation seems unlikely to improve in the near future.

Although these models focus on the antecedents of IRD, social identity and intergroup emotions theorists propose that similar appraisals of the *intergroup* situation shape GRD (Mummendey, Kessler, Klink & Mielke, 1999; Smith, Cronin & Kessler, 2008; Walker & Pettigrew, 1984). Social identity research shows that people will experience GRD if the intergroup situation is viewed as illegitimate, unlikely to improve without collective challenge and group boundaries are seen as impermeable (Mummendey et al, 1999).

Whereas social psychological RD research has focused on intergroup and interpersonal (upward and contrasting) social comparisons, political science RD research has focused on people’s comparisons with past, future, desired and deserved selves (e.g., Gurr, 1970). When the current situation violates expectations created by past experiences, people are more prone to feel politically alienated and to participate in collective protests (Gurr, 1970). More recent research extends this approach to include people’s comparisons of their ingroups at different points in time. Surveys of Mongolian and Russian citizens show that comparisons of one’s group to the group’s position in the past or future predicts levels of national identification (de la Sablonniere, Taylor, Perozzo & Sadykva, 2009; de la Sablonniere, Tougas & Lortie-Lussier, 2000).

**The Gap Between Measurement and Theory**

If RD is such a model social psychological theory with successful applications in a variety of important areas, why has it not been more generally effective? We believe the answer to this query involves the methods employed in many RD investigations. More precisely, we maintain that there are two major discrepancies between RD theory and common empirical operationalizations.

First, RD requires that people compare their situation to another possibility using principles about what “ought to be.” It is this emphasis on entitlement or “deservingness” that
distinguishes RD from other psychological theories that hold that people’s hedonic reactions occur through the comparison of experiences to particular referent points (e.g., adaptation-level theory, Helson, 1964; anchoring and adjustment, Kahneman & Tversky, 1982; prospect theory, Kahneman, 1992). For example, if people believe negative discrepancies are deserved or legitimate, they are unlikely to feel resentful, even though they recognize their disadvantage (Ellemers, 2002; Feather, 1999; Folger, 1987). Yet many researchers use perceptions of the magnitude of differences between one’s own situation and a relative standard to indicate RD, and neglect emotional reactions to those differences altogether. In contrast, RD theorists (Crosby, 1976; Folger, 1987; Martin, 1986a; Walker & Pettigrew, 1984) specify anger and resentment as essential affective consequences of RD. Thus, RD theory is basically an affective theory. When people feel they have been unfairly deprived relative to a meaningful comparison, they are held to be resentful and angry. And this affective response is predicted to mediate particular reactions in a host of domains.

Two different research traditions, both included in our meta-analyses, illustrate the role of affect in RD. The first tradition represents largely political science research that focuses on people’s comparisons with themselves at different points in time (e.g., Davies, 1962; Feierabend, Feierabend & Nesvold, 1969; Gurr, 1970. When the current situation violates expectations created by past experiences, people are held to feel politically alienated and more likely to participate in collective protests (e.g., Gurr, 1970; Herring, 1989, Newton, Mann & Geary, 1980).

Yet researchers in this tradition often use measures of aspirations as a proxy for violated expectations (Gurney & Tierney, 1982). For example, a popular method for measuring RD, the Cantril-Kilpatrick Self-Anchoraging Scale (Cantril, 1965), has respondents place themselves on a 10-step ladder with the top rung labeled as the best possible life and the bottom rung as the worst possible life. When defined in this way, this scale measures discrepancies between people’s
attainments and aspirations; but it does not measure discrepancies between their expectations as to what they deserve and their current situation (Finkel & Rule, 1987). Thus, this measure emphasizes RD’s cognitive component at the expense of its crucially important affective component.

Even researchers who do not use the Cantril-Kilpatrick Scale often blur the difference between what is considered just, satisfying or expected (e.g., Martin, 1986b). Although researchers may not make these distinctions, research participants do. When employees at a manufacturing company were asked to create two pay plans, one based on what they expect and the other on what is just, they produced significantly different distributions (Martin, 1986b). We believe that studies that fail to take these distinctions seriously obscure the role of justice-related affect and are not testing RD theory. Thus, with our “affect hypothesis,” we predict that:

Hypothesis 1. *RD measures that tap either affective judgments or both affective and cognitive judgments will be more strongly related to key outcome variables in comparison to purely cognitive measures that ask only for respondents to estimate differences between their present situation and a referent comparison.*

Second, if people do not make the relevant subjective comparison; they should not experience RD – a requirement that distinguishes RD from the earlier frustration-aggression hypothesis and other models of social justice and discrimination. Yet researchers often construct RD measures from objective demographic characteristics without explicit comparison referents. For example, Boyce, Brown & Moore (2010) draw upon a representative longitudinal sample of British households to estimate participants’ life satisfaction. They compared participants’ self-reported income to the income of reference groups constructed by averaging the incomes of participants in similar geographic regions, age groups and education levels to show that one’s relative rank with an income reference group predict life satisfaction. Although these patterns are
instructive, we still do not know whether participants made or reacted to the same comparisons that researchers constructed from various geographic, gender and occupational groups. Indeed, such a comparison seems highly unlikely. Such findings are interesting in their own domain, but they are not testing RD theory.

More importantly, the level of analysis represented by the RD comparison must match the level of analysis represented by the outcome. Runciman’s (1966) basic conceptual distinction contrasts individually-based “egoistic” RD (IRD) produced by interpersonal comparisons and group-based “fraternal” RD (GRD) produced by intergroup comparisons. Repeated research demonstrates that it is feelings of group RD that promote support for political protest (Pettigrew, 1967; Walker & Mann, 1987) and outgroup prejudice (Pettigrew & Meertens, 1995; Pettigrew et al., 2008; Vanneman & Pettigrew, 1972). Yet researchers often ignore this distinction (Walker & Pettigrew, 1984) – for example, by using interpersonal comparisons to predict collective behavior (e.g., Long, 1975; Newton, Mann & Geary, 1980; Useem, 1975). This confusion of levels may explain many of the literature’s negative RD results. Feeling deprived may inspire participation in collective behavior, but only if the person feels deprived on behalf of a relevant reference group. Similarly, IRD should predict individual behavior (e.g., “moonlighting” to earn extra money, stealing, using drugs). Unfortunately, previous RD reviews have either excluded (Crosby, 1976; Cook, Crosby & Hennigen, 1977) or ignored (Finkel & Rule, 1987; McPhail, 1971) research that distinguishes between IRD and GRD. Therefore, we suspect the conclusions of some earlier literature reviews are unduly pessimistic (e.g., Finkel & Rule, 1987; Guerney & Tierney, 1982; McPhail, 1971).

The recognition of two patterns of RD also raises a subtle distinction between levels of analysis that are easily confused empirically (Kawakami & Dion, 1993; Smith, Spears & Oyen, 1994). Researchers often ask subjects to compare themselves to a single member of another
group. Suppose we ask a woman to compare her personal working situation with those of male employees. On the one hand, the comparison to male employees may indicate an intergroup comparison with gender as the salient comparison. The employee sees herself as a representative female employee in comparison to male employees. On the other hand, the emphasis on her personal working situation may indicate an interpersonal comparison with her unique individual characteristics salient. The employee sees herself as an employee in comparison to another employee in the same company. Thus, the same comparison could represent either an intergroup comparison and measure GRD, or an interpersonal comparison and measure IRD. Indeed, it could even involve both types of comparisons.

Such ambiguity suggests that GRD will only be associated with collective behavior if people view themselves as group representatives in comparison to an outgroup. Therefore, we must distinguish between three types of social comparisons: those between (1) oneself and an ingroup member, (2) oneself and an outgroup member, and (3) one’s ingroup and outgroup.

From these considerations, we propose a “fit hypothesis”:

Hypothesis 2. *The relationship between RD and various dependent measures will be stronger when the level of reference for both the RD and outcome measures is the same.*

Third, there is the issue of measurement quality. If there is a solid relationship between RD and an outcome variable, then studies with well-measured RD and outcome variables should record stronger effect sizes then studies with less rigorously measured variables. Pettigrew and Tropp (2006) found this to be true in their meta-analysis of intergroup contact effects – thus lending support to the strength of the relationship between contact and reduced prejudice.

Hypothesis 3. *The relationship between RD and various outcome measures will be stronger when higher quality measures are used for the RD and outcome measures.*
Reactions to Relative Deprivation: The Outcome Variables

Although reviews of RD research center largely on whether RD offers an adequate explanation for participation in collective action (e.g., Brush, 1996; Finkel & Rule, 1987; Guerney & Teirney, 1982), collective action is just one of many possible responses to feeling deprived. Crosby (1976) lists four reactions to IRD: (1) stress symptoms, (2) self-improvement, (3) violence against society and (4) constructive change of society. Mark and Folger (1984) outline three types of responses to RD: (1) attitudes and behavior toward the self, (2) attitudes and behavior toward the comparison object that triggers IRD, and (3) attitudes and behavior toward the system. Finally, Wright (1997) lists five types of reactions to perceived injustice (including injustices created by disadvantageous comparisons): (1) inaction, (2) individual normative actions, (3) individual non-normative actions, (4) collective normative actions, and (5) collective non-normative actions.

Together, these theoretical analyses make two critical distinctions. First, RD reactions can represent intentions and behaviors to improve one’s personal situation, or they can represent intentions and behaviors to improve the situation for one’s reference group generally (Ellemers, 2002; LaLonde & Silverman, 1994; Wright, 1997). Second, these frameworks distinguish among internal states (such as anxiety, depression and attitudes toward the self), attitudes (towards the system and other groups) and actual behavior.

We draw upon these two distinctions to organize the RD literature into four classes of outcome variables. The first category of research includes collective behavior, the primary focus of early RD work and the recipient of the most severe criticisms. Based on our fit hypothesis, we predict that GRD will be the most closely related to collective action. The second category entails studies of intergroup attitudes that include attitudes toward such political policies as affirmative action and immigration, prejudice toward outgroups and ingroup identification and
bias. Here we predict such attitudes also tap the group level of analysis, but they do not refer to collective action. The third category of research includes *individually oriented behaviors*, both normative and non-normative. We predict that IRD will be the most closely related to individual behavior. Finally, the fourth category entails work on such *internal responses* as psychological stress, depression, physical health and altered self-evaluations. The fit between internal states and different RD comparisons is less clear. One could argue that IRD will be most closely related to internal states (particularly if those states are defined as representing the unique personal self).

**Method**

**Inclusion criteria**

We found that much of the research literature that claims to be testing RD theory does not do so in clear-cut, direct ways. Clearing this underbrush became an essential first step. This made our task of designing strict inclusion rules – a key requirement of meta-analysis – especially important for this study. We employ six inclusion criteria for the RD measures and two inclusion criteria for the outcome measures. Below we make explicit what types of studies are being excluded by these criteria, as recommended by Rojahn and Pettigrew (1992).

*Criterion 1.* Because we focus on the relationship between feelings of deprivation and behavior, we consider *only those empirical studies in which researchers treat subjective RD measures as a predictive variable* (excluding 146 studies that we reviewed). This requirement excludes most experimental investigations designed to investigate the antecedent conditions of RD (e.g., Bernstein & Crosby, 1980; Folger & Martin, 1984; Olson & Ross, 1986). In the excluded experimental studies, situational variables are manipulated to create feelings of deprivation, then subsequent evaluations of fairness and satisfaction are measured (e.g., Cooper & Brehm, 1972; Martin, 1986a; Spector, 1956).
This criterion also excludes case studies in which researchers used RD as an organizing principle for their analysis. These qualitative analyses ranged from analyses of inequities in transportation access for women in Northeast England (Dobbs, 2007) to understanding the exodus of the upper-middle class elite from Buganda to other parts of East Africa (Thompson, 1989).

Criterion 2. **Relative deprivation or a close proxy variable must be asked of individual subjects directly.** This decision excludes 99 investigations in which researchers infer feelings of deprivation via aggregate measures of such variables as income inequality. For example, Gurr measured economic RD with (a) short term trends in the amount of exports and imports, (b) cost of living increases, (c) GNP growth rates and (d) summaries of news stories about crop failures, unemployment and other adverse economic conditions (Gurr, 1970; Gurr & Duvall, 1973). Similarly, Fierabend and colleagues (1969) measured systemic frustration by combining a country's gross national product (GNP), the caloric intake per capita, the number of physicians, telephone, newspapers and radios per unit of population, degree of literacy and urbanization.

These studies commit the ubiquitous ecological fallacy. This fallacy draws conclusions about individuals from macro-level data alone – a mistake often seen in statements made about individual voters from aggregate voting results (Pettigrew, 2001). It is a fallacy because macro-units are too broad to determine individual data, and individuals have unique properties that cannot be safely inferred from macro-level data alone.

This requirement also excludes 202 studies in which RD is a state inferred from demographic variables collected from individual respondents (e.g., Geschwender & Geschwender, 1973; Pinard, Kirk & Van Eschen, 1969; Reagan, Salsberry & Olsen, 2007; Siahpush, Borland, Taylor, Singh, Ansara, & Serraglio, 2006). Although these investigations measured characteristics at the individual level, there is no evidence their subjects interpreted their personal demographic
characteristics as hypothesized by the researchers. For example, Orum and Orum (1968) defined students as relatively deprived if a demographic measure of their parents’ educational status was inconsistent with a demographic measure of their parents’ socio-economic status.

Among the excluded studies is research in which relative deprivation is defined as a hierarchical linear model interaction effect in which adolescents’ assessments of their economic deprivation are more predictive of their individual attitudes and behavior the less economically deprived the local neighborhood (Bergburg, Thorlindsson & Sigfusdottir, 2009). The authors argue that adolescents in this situation are more likely to make unfavorable social comparisons. Epidemiologists also uncover relationships between relative income and (1) migration patterns (Oded & Edward-Taylor, 1999) and (2) mortality rates and physical health (e.g., Reagan et al, 2007; Siahpush et al, 2006) by constructing comparisons between respondents’ income and the incomes of other respondents from the same census tract (e.g., D’Ambrosio & Frick, 2007, Eibner, Sturm, & Gresenz, 2004). Interesting as these studies are, they are not tests of RD.

We also exclude experimental studies in which researchers manipulate RD as a between subjects variable (e.g., Markovsky, 1988; Ross & McMillan, 1973; Taylor, Moghaddam, Gamble & Zeller, 1986). For example, Mark (1984) operationalized RD as the relative likelihood told to subjects as to whether they could participate in the study. Grant (1992) operationalized RD as (fictitious) male judges' rejection for sexist reasons of an opinion paper written by a group of female subjects. In this research, researchers did not ask individual subjects how satisfied or frustrated they were in comparison to some standard.

Criterion 3. For two reasons, we exclude 94 studies in which investigators created difference scores from respondents’ answers to different survey questions to represent participants’ subjective experience (e.g., Buunk, Zurriaga, Gonzalez-Roma, & Subirats, 2003; Taylor, 2002; van Dierendonck, Schaufeli & Buunk, 2001). First, difference scores are inherently
difficult to analyze and interpret (Pedhazur & Schmelkin, 1991: 291-295). Second, difference scores do not insure that the respondent made the same comparison used by the investigator.

**Criterion 4. Researchers must operationalize RD as a comparative construct** (excluding 25 studies that we reviewed). This requirement excludes studies in which researchers defined RD as respondents’ feelings of injustice or resentment about their general personal or group situation (Dube & Guimond, 1986, study #3; Van Kyk & Niewoudt, 1990). Although Appelgryn & Nieuwoudt (1987), for example, asked white Afrikaners and black South Africans to rate whether their personal and group political, social and economic situations were just, the respondents did not make these ratings in comparison to any particular standard or referent. The comparative requirement also excludes simple counts of the number of incidents in which a person experiences discrimination (Dion, 1986) and scales theoretically linked to RD concepts but explicitly designed to measure political alienation (e.g., Kluegal & Smith, 1986; Simmel, 1961). None of these measures specify clear comparison referents.

**Criterion 5. We only include RD measures if they were negative discrepancies that created feelings of deprivation** (excluding 6 studies that we reviewed). This criterion excludes measures that tap what Beaton and Deveau (2005) and others define as RD on the behalf of others or ideological deprivation. For example, Tougas and Veilleux (1990) measured men's perceptions of differences in salary, hiring and promotions between men and women multiplied by their dissatisfaction with these discrepancies. Unlike the traditional conceptualization of GRD, this measure assumed that the greater the (illegitimate) discrepancies favoring their group, the more dissatisfaction and resentment men felt. (The authors also distinguished this measure from a measure of collective deprivation that captured the extent to which men felt men were deprived in comparison to women).
Criterion 6. We only include RD measures if the relationship between the respondent and the comparison target is clear (excluding 15 studies that we reviewed). For example, if a RD measure represents a comparison between two outgroups, we exclude the measure. Thus, we eliminate African-American respondents' comparisons between well-educated blacks and blue or white collar workers (Abeles, 1976). This is necessary because it is unclear whether these African-American respondents viewed well-educated Blacks as part of their ingroup or as an outgroup. If a RD measure is a comparison between a subgroup and the whole group, we exclude the measure for the same reason. So we could not use unemployed Australians' comparisons between their immediate peer group and all unemployed Australians (Walker & Mann, 1987). Unemployed respondents may feel equally identified with both their unemployed peers and all unemployed Australians, or may see both groups as essentially the same. Finally, we exclude the intriguing measures of comparative mistreatment used by Guyll, Matthews and Bromberger (2001) to uncover differences in cardiovascular reactivity for African-American and European-American women. It was unclear in this study just who respondents had in mind when they reported that other people had been treated with more respect, courtesy and better service (e.g., did these other people represent an ingroup or outgroup?).

We also exclude studies in which RD measures were the product of several different comparisons (Ashton, 1978; Dibble, 1981; Vanneman & Pettigrew, 1972). For example, Ashton (1978) compares respondents who report feeling deprived compared to the outgroup and the ingroup ("doubly deprived" respondents) with respondents who report feeling deprived compared to the outgroup but gratified compared to the ingroup ("group deprived" respondents). Unfortunately, there is no way that we could disentangle these comparisons in order to test the fit hypothesis.
Moreover, we exclude studies in which researchers asked people about the source of their comparison information without an evaluation of the outcome or asked participants to indicate the frequency with which they made a particular type of comparison (e.g., Brown, Ferris, Heller & Keeping, 2007). Without clear evidence as to how people reacted to a specific comparison, we could not be sure that they had experienced RD (see also Stephan, et al, 2002).

In addition, we require each outcome measure to meet two inclusion criteria.

**Criterion 1.** We include data only if the relationship between the outcome measure and the respondents' attitudes and behavior is clear (excluding 4 studies that we reviewed). Sears and McConahay (1970) included a measure of the number of riot acts that respondents witnessed. Because it is unclear how the observation of collective behavior relates to feeling motivated to act from feelings of deprivation (one could just be in the "wrong" place at the "wrong" time), we exclude these data. We also exclude opinions about whether a non-membership group should use violence to accomplish political goals (e.g., whether whites thought blacks should use violence). Finally, we exclude measures in which respondents reported what they thought other peers or family might think of them (Brevik & Olweus, 2006; Brunsting & Postmes, 2002). Although RD might be related to perceptions of one’s neighborhood as less friendly, it is unclear how feeling deprived would directly cause or predict neighborhood climate.

**Criterion 2.** We only include data if the outcome measure is a measure that has not been used by other researchers as a RD measure (excluding 28 studies that we reviewed). For example, some authors consider awareness of a negative discrepancy as a measure of RD; but others suggest that it is the discrepancy and the resulting affective consequences that together produce RD. So we excluded studies in which researchers defined RD as the awareness of differences in a particular domain and the dependent measure as the resulting feelings about these differences. This choice meant the exclusion of research in which researchers used relative
differences or discrepancies in material goods or income to predict income satisfaction (Crawford Solberg, Diener, Wirtz, Locus & Oishi, 2002; Johnson & Johnson, 1999). Similarly, we exclude projects in which researchers used participants' ratings of the fairness of a particular comparison to predict their satisfaction with the comparison (Austin, McGinn & Susmilch, 1980; Messe & Watts, 1983).

Measures of perceived discrimination illustrate the same difficulty. Some researchers use measures of perceived discrimination as a measure of RD (Dibble, 1981; Dion, 1986); others treat perceived discrimination as a result of feeling (group) deprived (Kommen & Frankel, 1992; Long, 1975). Because we cannot consider the same construct as both an independent and dependent variable, we removed studies with perceived discrimination as an outcome measure.

**Locating Relevant Studies**

The selected studies were part of more than 860 studies located by using a variety of methods: (1) a computer search through psychological, sociological, economic, political and dissertation abstracts through January 2010; (2) 112 personal letters and e-mails to researchers who have published relevant studies; (3) a review of reference lists from previously located studies and conference presentations; and (4) “list serve” requests to members of the International Society for Justice Research, Society for Personality and Social Psychology, Society for the Study of Social Issues, International Society for Political Psychology, the European Association of Social Psychologists and the Society of Australasian Social Psychologists. Key words for the data base search included: (employment) over qualification, subjective social status, upward (contrasting) social comparisons, injustice gap, referential comparisons, perceived socio-economic differences, relative deprivation, relative disadvantage and relative discrimination. We also examined the reference sections of all located primary studies and relevant literature reviews.
The search yielded 210 studies (summarizing 293 independent samples and data from 186,073 participants) written between 1961 and January 2010 that met our inclusion criteria (median year of publication = 2000). Note the sharp decline from the 860 studies initially obtained to 210 final studies. This large 76% drop-off represents the loss from our necessary criteria for inclusion. This massive clearing of the underbrush demonstrates that most of the research literature labeled as testing RD does not actually test RD theory directly. Indeed, this has been a major problem for previous qualitative reviews that did not use strict inclusion criteria.

Most of the included studies were published journal articles (168), but the data set also includes eight book chapters and 34 unpublished studies (including 19 dissertations). Samples from the United States represent 79 (39%) of the studies, but 29 other countries are also represented. Similarly, social psychology represents the most frequent author and journal affiliation (45%). However, the data set also includes work published in other sub-disciplines of psychology (e.g., personality, political, education, industrial/organizational, clinical/counseling, economic and health psychology) and disciplines outside of psychology (e.g., criminology, textile arts, economics, public health, political science, sociology and behavioral medicine). Although almost all the studies were written in English, the final data set also includes studies written in French, German and Afrikaans. Samples ranged from probability population surveys to single occupations (e.g., university faculty, female police officers, funeral directors and concrete construction workers) to ethnic, religious, national and political minority and majority groups.

It is important to show that these inclusion criteria do not inflate our estimates of average effect sizes. Actually, estimates of effect sizes for three categories of excluded studies suggest that our criteria lead to a more conservative test of our hypotheses. For example, the average mean effect size for the 20 independent samples that we excluded because the outcome domain was too similar to the RD measure domain was +.32 (ranging from +.23 to +.40). For the 32
excluded tests of the relationship between RD, manipulated as a between subjects' variable, and subsequent attitudes and behaviors, the average effect size was +.21 (ranging from +.17 to +.25). For 23 excluded tests of the relationship between aggregate societal level characteristics and outcome measures, the average effect size was +.51 (ranging from .40 to .61).

In contrast, as we describe below, the average effect size for the 210 separate studies included in the final data set was +.11, strongly indicating that our exclusion rules constitute a quite conservative test of the relationship between RD and various dependent measures. However, the choice to eliminate researcher-created difference scores as RD measures had scant effect. For 33 excluded studies that used difference scores to measure RD, the average effect size was +.09 (ranging from +.03 to +.16). This finding is consistent with our prior contention that researcher-created difference scores do not adequately capture RD (see also Smith & Ortiz, 2001).

**Variables Coded From Each Study**

We recorded the following general information (if available) from each study: (a) date of publication, (b) publication form, (c) author and publication discipline, (d) sample nationality, (e) number of respondents, (f) respondents’ age, (g) respondents’ gender, (h) type of sampling strategy and (i) response rate. For each study, if the sample could be considered homogeneous, we also coded the relevant group membership and occupation. Two coders independently read and coded each sample, RD and outcome measure. All coding was done independently of effect size calculations, and any disagreements were solved through discussion between the two coders. The reliabilities for coding were high (all kappas above .90).

*RD measures.* For each different measure of relative deprivation, we coded whether participants estimated a difference (a cognitive measure), reported how they felt about the difference (an affective measure) or indicated whether their relative disadvantage was undeserved
or unfair (treated as a second form of an affective measure). If measures of mood or emotions were woven into the RD measure, we coded it as an affective measure.

We also coded whether the comparison was between (1) the respondent's personal situation and the situation for a member of their ingroup, (2) the respondent’s personal situation and the situation for a member of an outgroup, (3) their ingroup's situation and an outgroup's situation or (4) the respondent’s past or present experience with their current situation, future expectations or theoretical possibilities (e.g., the best possible life). If a measure included comparisons with both an ingroup and an outgroup referent (e.g., questions including a female and male employee referent for female employees, Hafer and Olson, 1992), the measure was coded as representing an outgroup comparison.

We also coded whether the comparison dimension represented differences in (1) social position, income or class, (2) housing or income, (3) standard of living or purchase ability, (4) wages or taxes, (5) academic or task performance, (6) work situation, (7) geographic location, (8) health or overall well-being, (9) general life situation, (10) political power or institutional treatment, (11) physical appearance or personality characteristics or (12) one’s relationships (e.g., from a single parent family in comparison to a two parent family, Brevik & Olweus, 2006).

Finally, to test our hypothesis 3, we coded the quality of each RD measure. The first category included single-item measures and was assigned a score of one. The second category consisted of multiple-item scales with unknown reliability or an alpha of less than .70 and was assigned a score of two. The third category consisted of multiple-item scales with an alpha of .70 or more and was assigned a score of three. If at least one item in a scale met our inclusion criteria, we included it as part of the database. We also distinguished between measures that had been used in a previous research project (e.g., the subjective status measure - Goodman, Adler, Kawachi, Frazier, Huang & Colditz, 2001) or created for the current project.
Outcome measures. We classified each dependent measure represented in the larger data set into four general categories. The first category contains outcomes that we describe as internal states; it includes: (1) stress, anxiety, depression, hopelessness, mental illness and pessimism, (2) (personal) self-esteem and self-efficacy and (3) (poor) physical health (e.g., more obesity, heart disease, restricted sleep). The second category contains outcomes that we term individual behavior; it includes: (1) forms of deviance (e.g., violence, stealing, and counterproductive work behavior), (2) forms of escape (e.g., smoking, drinking, drug use, absenteeism and social isolation) and (3) forms of achievement (e.g., moonlighting, academic performance). The third category involves intergroup attitudes; it includes: (1) attitudes toward the ingroup (e.g., ingroup favoritism, nationalism and identification), (2) attitudes toward the system (e.g., voting intentions, support for authorities) and (3) attitudes toward outgroups (including prejudice, political attitudes toward immigration and affirmative action). Finally, the fourth category contains outcomes that entail collective behavior; it includes: unstructured and structured forms of collective action and approval of political violence.

We coded each outcome measure as representing an attitude, a behavioral intention or an actual behavior. We also coded the quality of the outcome measure with the same three measurement quality categories listed above. There was an additional category for outcome measures defined as difference scores created by the researchers. We also distinguished between measures that had been used in previous research or created for the current project.

Computation and Analysis of Effect Sizes

We report Pearson’s $r$ as the principal indicator of effect size throughout the analysis (Rosenthal, 1995). All mean $r$’s were computed with each effect size weighted by the reciprocal of its variance (which gives more weight to effect sizes that are more reliably estimated, see Borenstein, Hedges, Higgins & Rothstein, 2009). Although the use of random effects analysis
addresses the issue of sampling error, we cannot correct for measurement error because of the failure of most RD studies to supply the needed data. As Schmidt (2010) emphasizes, this means that our final effect sizes are likely to be smaller than if we were able to eliminate measurement error.

A positive mean effect size indicates that greater RD relates to more of the particular behavior or stronger attitudes. If no correlations were reported (as was the case for 11.5% of the included effects), the effect size was derived from the results of significance tests (chi-squares, t or F ratios) by use of the conversion formulas provided by Johnson (1993). If a particular relationship was reported as non-significant or the result was completely omitted (but implied by the methods section as was the case for two effect sizes), we assign a value of .00 for the effect size. This procedure, as Rosenthal (1995) points out, is a conservative one that reduces the mean effect size.

For three reasons, our primary unit of analysis is the individual test of the relationship between a single RD measure and a dependent measure. First, many studies include separate questions for different comparison targets. Given our expectation that different types of comparisons will be more or less closely related to different outcomes, these tests had to be treated separately. Similarly, researchers often include separate cognitive and affective measures of RD within the same study. Finally, many studies include several different outcome measures. The mean number of tests included for a single study was 8.51 (median = 4, ranging from 1 to 89 tests per study). However, studies that included larger numbers of tests also reported smaller average effect sizes ($r (208) = -.16, p<.05$), indicating that, if anything, a focus on the effect sizes of individual tests is a conservative method for estimating the strength of RD effects. All combined effect sizes were calculated using Rosenthal's (1995) suggested formula.
For each category of effect sizes, we calculate the weighted mean effect size and the corresponding 95% confidence interval. Second, we examine the homogeneity of each set of effect sizes by calculating the homogeneity statistic Q that has an approximate chi square distribution with k-1 degrees of freedom, where k is the number of effect sizes (Borenstein et al, 2009). In the absence of homogeneity, we test a series of categorical models that relate the effect sizes to characteristics of the study (Borenstein et al, 2009). We use a random effects model for all our analyses, because we assume that the variance around the mean effect size cannot be fully explained by potential moderators due to the heterogeneity of our sample (Borenstein et al, 2009). This approach is a more conservative test of our hypotheses in comparison to a fixed effects model analysis. The weighted averages of the effect sizes are computed on the basis of the Fisher Z transformed correlation by using the inverse of the variance as the weights. The reliability for mean effect sizes was established through calculations of confidence intervals as well as Stouffer’s Z.
Results

Table 1 lists the mean effect size for each of the three levels of analysis (studies, samples and individual tests) and the four types of outcome variables. Each of the mean effects is highly reliable, but relatively small – ranging from $r = +.11$ to $+.17$. As shown in Table 1, the studies provide smaller effects on average in comparison to independent samples and individual tests. Similarly, tests for internal states yield significantly larger mean effects that either those for individual behavior ($Q_B = 11.40, p < .001$) or intergroup attitudes ($Q_B = 14.08, p < .01$), but not reliably larger in comparison to collective behavior ($Q_B = 1.41, p = .24$). Tests of collective behavior yield slightly larger mean effects in comparison to intergroup attitudes ($Q_B = 2.87, p = .09$) but not compared to individual behavior ($Q_B = 2.25, p = .13$). Finally, there is no reliable difference between individual behavior and intergroup attitude tests ($Q_B = 0.03, p = .84$). Note, however, the considerable heterogeneity within each category ($Q_w$ ranged from 988.20 to 4,090.63) before the application of the corrections advanced by our hypotheses.

Publication bias

Before we test our hypotheses, we must check for potential publication biases. First, we applied Rosenthal’s (1995) failsafe index. For 293 samples, it would require more than 8,111 missing samples that reported an effect size of zero to erase the statistical significance of the mean RD effect size at the 5% confidence level. This number is considerably larger than the 860 studies uncovered by our intensive literature search. Second, we determined that the relationship between sample sizes and effect sizes is not statistically significant ($r (292) = -.058, p = .47$). In other words, there is no evidence that smaller sample sizes are contributing larger effect sizes (one potential determination of whether a study is published). Third, Figure 1 provides a scatter diagram comparing sample sizes with Fischer’s Z representing the effect sizes. The graph roughly resembles a funnel. The funnel is not sharply skewed and the mean effect remains approximately
the same regardless of the sample sizes. This graph also suggests that no strong publication bias is operating in these data.

Finally, as a direct test for publication bias, we compare the mean effect sizes from different publication sources. The nine book chapters yield the strongest average effect size \( r = +.15, CI = -.02 \text{ to } +.27, n = 4,589 \), the 167 published journal articles yield the next strongest average effect size \( r = +.11, CI = +.09 \text{ to } +.14, n = 123,772 \), the 19 dissertations yield a smaller average effect size \( r = +.06, CI = -.01 \text{ to } +.13, n = 4,305 \) and the 15 unpublished papers yield the smallest average effect size \( r = +.05, CI = -.06 \text{ to } +.12, n = 10,252 \). An unfocused test of a between classes effect for the four types is not statistically significant \( Q_B(3) =3.55, p = .32 \); and within each publication category, there was substantial heterogeneity among effect sizes \( Q_W \).

Testing the Affect Hypothesis

Table 2 combines all our data across the four categories of outcome variables to test the affective hypothesis. For all three levels of analysis, our first hypothesis is supported. The mean effects for studies, samples and tests that employ affective RD measures are significantly larger than those employing purely cognitive RD measures. This special strength of affective RD measures was largely found among measures of individual behavior \( r \) (cognitive RD) = +.07 vs. \( r \) (affective RD) = +.16, \( Q_B(1) = 13.61, p < .05 \) and intergroup attitudes \( r \) (cognitive RD) = +.10 vs. \( r \) (affective RD) = +.15, \( Q_B(1) = 4.24, p < .05 \). The difference between affective and cognitive RD measures for internal states did not attain statistical significance \( r \) (cognitive RD) = +.16 vs. \( r \) (affective RD) = +.20, \( Q_B(1) = 3.2, p = .97 \). Nor was there a statistically significant difference
between affective and cognitive RD measures for collective behavior \( (r \text{ (cognitive RD)} = +.15 \text{ vs. } r \text{ (affective RD)} = +.14, Q_B(1) = 0.05, p= .82) \). However, within all categories of affective RD measures, there was significant heterogeneity among effect sizes \( (Q_w \text{ ranged } 364.46 \text{ to } 1,818.65) \).

**Testing the Fit Hypothesis**

The second hypothesis is that RD measures will predict outcome measures more strongly if the level of analysis represented by the RD measure matches the outcome measure level. Before testing this possibility, we excluded RD measures in which the respondent compared themselves to an outgroup member. As discussed earlier, it is unclear whether the respondents view this type of comparison as interpersonal (between themselves and another unique person) or as intergroup (between themselves as a representative of their ingroup and the outgroup). We also excluded outcome measures that focused on attitudes toward one’s ingroup because we could make no clear theoretical predictions. Depending upon respondents’ attributions about the cause or stability of RD, increased (individual or group) RD could lead to more or less identification or commitment to one’s ingroup. Alternatively, ingroup identification may be best viewed as a moderator of the relationship between RD and other outcomes (e.g., more identified respondents will react more strongly to group RD; see van Zomeren, Postmes & Spears, 2008).

We hypothesized that group-based comparisons (to other groups or to one’s group at another point in time) would more strongly predict collective behavior, attitudes toward the outgroup and the larger system. We further hypothesized that individual-based comparisons (to other ingroup members or one’s self at another point in time) would more strongly predict individual achievement, deviant behavior as well as mental and physical health. If the comparison and outcome represent the same level of analysis (e.g., intergroup RD comparisons associated with collective behavior), we coded it as a “match”. If the comparison and outcome clearly represent different levels of analysis, we coded it as a “mismatch” (e.g., an intergroup comparison
associated with individual academic achievement). As shown in Table 3, matched tests outperformed mismatched tests whether we consider all relevant RD comparisons or limit the analysis to affective RD measures. If we compare matched tests in which we screen for quality (by excluding single-item measures, outcomes measured with difference scores and convenience samples), matched tests yield an effect size of +.23 ($CI = +.17$ to $+.29$, $k=43$, $n = 27,064$) in comparison to mismatched tests ($r = +.11$, $CI = +.01$ to $+.21$), $k = 27$, $n = 16,453$). This pattern remains the same if we limit the analysis to comparisons to other people or groups (as assumed by the original Runciman hypothesis). Thus, the results in Table 3 support all three of our hypotheses.

Comparisons to outgroup members

The choice to exclude comparisons to outgroup members from the tests described above may mask the potential importance of these comparisons. For example, the six affective comparisons between the respondent and an outgroup member yielded strong effect sizes for personal self-esteem and efficacy ($r=+.31$, $CI = +.24$ to $+.39$); the 18 affective comparisons between the respondent and an ingroup member ($r=+.13$, $CI = +.04$ to $+.29$) yielded a much smaller effect size ($Q_B(1) = 8.35$, $p<.05$). The six tests of outgroup member comparisons for mental and physical health yielded an effect size of +.28 ($CI = +.19$ to $+.36$) that was similar to the 13 tests of comparisons to ingroup members ($r=+.29$, $CI = +.06$ to $+.48$; $Q_B(1) = 0.01$, $p=.94$). The eight affective tests predicted the desire to escape or exit the situation most strongly ($r = +.23$, $CI = +.14$ to $+.32$); twice the size of the effect sizes yield by 13 tests for comparisons to ingroup members ($r = +.11$, $CI = +.04$ to $+.17$; $Q_B(1) = 4.44$, $p<.05$). However, ten affective tests of these comparisons to outgroup members yielded smaller effect sizes for collective behavior ($r = +.16$, $CI = +.12$ to $+.21$) in comparison to 19 affective intergroup comparisons ($r = +.20$, $CI = +.14$ to $+.26$; $Q_B(1)=.88$, $p=.35$). Finally, all three types of affective social comparisons - to an
ingroup member ($r = +.23, CI = +.15 \text{ to } +.31, k = 7$), to an outgroup member ($r = +.23, CI = +.09 \text{ to } +.37, k = 3$) or an intergroup comparison ($r = +.23, CI = +.08 \text{ to } +.39, k = 9$) - were associated with negative attitudes toward the larger system.

**Testing the Research Quality Hypothesis**

We compared 358 single-item RD measures to 122 multiple-item RD measures with clear reliability and 233 multiple-items with unknown or poor reliability. Single-item measures predict outcomes ($r = +.12, CI = +.10 \text{ to } +.15$) as strongly as multiple-items with poor reliability ($r = +.12, CI = +.09 \text{ to } +.21$) but significantly less strongly in comparison to multiple-item measures with solid reliability ($r = +.18, CI = +.15 \text{ to } +.20$: $Q_{df}(2) = 12.82, p < .05$). A direct comparison between single item and all multiple item measures showed a reliable difference ($Q_{df}(1) = 9.53, p < .05$). There remained, however, significant heterogeneity within each category ($Q_w$ ranged from 1,220.52 to 4,030.42).

Next, we compared 159 single-item outcome measures to 322 multiple-item outcome measures with solid reliability, 187 multiple outcome measures with unknown or poor reliability and 23 difference score outcomes. Difference scores yielded the weakest effects ($r = +.02, CI = -.02 \text{ to } +.06$), single-items yielded somewhat stronger effects ($r = +.12, CI = +.10 \text{ to } +.14$), and multiple-item scales yielded stronger effects (for solid reliability measures - $r = +.14, CI = +.13 \text{ to } +.16$; and for those with unclear reliability measures - $r = +.15, CI = +.13 \text{ to } +.17$; $Q_{df}(3) = 33.95, p < .05$). Again, a direct comparison between all multiple item and single item measures showed a marginally reliable difference ($Q_{df}(1) = 3.45, p = .06$).

Multicollinearity could be a problem here, because affect relates significantly with RD measure quality ($r(712) = +.33, p < .05$), outcome measure quality ($r(712) = +.11, p < .05$), and matched levels of analysis ($r(495) = +.12, p < .05$). Thus, the affect effect may reflect largely its relationships with RD quality and matching levels of analysis. To test this possibility, we treated
effect size as the outcome in a modified weighted least square regression analysis (Lipsey & Wilson, 2001). Outcome quality ($\beta = .038$, $Z=4.50$, $p<.001$), RD quality ($\beta = .097$, $Z=11.49$, $p<.001$), affect ($\beta = .11$, $Z=7.54$, $p<.001$) and whether the levels of analysis were matched or mismatched ($\beta = .16$, $Z = 10.73$, $p<.001$) all reliably predicted effect size. The regression model was statistically significant ($Q(4) = 31.59$, $p<.001$) with a random effects variance component $v = .02$ and an explained variance of 5.4%.$^3$ As one might expect, more theoretically accurate RD measures represent better quality and are matched with better quality outcomes. But the important point is that theoretical accuracy (defined as the inclusion of justice-related affect and matched levels of analysis) continues to predict larger effect sizes.

Figure 2 summarizes our hypotheses-testing results. Using individual tests, the histogram shows that the combined predictive effect of relative deprivation for all tests is +.13. The mean effect size for those tests that have none of our three hypothesized improvements – that is, they are the worst subset of tests because they used only low-quality and cognitive measures of RD and related them to outcome variables at a different level of analysis – is only +.08. This effect rises to +.17 for those tests that employ some type of affective measure of relative deprivation. The effect for affective tests only rises once more to +.20 when the analysis level for relative deprivation is the same as that of the outcome variable. Finally, the mean effect reaches +.23 for the optimal subset of affective tests that boasted both a level fit and higher quality measures. In terms of variance explained, Figure 2 reveals that tests that employed all three of our methodological improvements were almost three times stronger that the mean effect for all the RD tests in our file and more than eight times stronger than the worst subset of tests that featured none of our improvements.

**Discussion**
Although the initial analysis suggests that the mean RD effect size might be as weak as some critics claim, if the RD measure includes affect and matches the level of analysis implied by the outcome variable, the predictive power improves sharply. The initial effect size of +.14 (for independent samples) improves to +.23 for high quality affective RD measures matched to the outcome level of analysis. These patterns emerged despite our strict inclusion rules and our file’s enormous variety of ages, nationalities and group memberships. More importantly, appropriately measured RD predicts a wide variety of important outcomes - including behavior, behavioral intentions and attitudes as well as measures of mental and physical health.

This mean effect size of +.23 is comparable to those of other meta-analyses of important social psychological processes. Thus, Pettigrew and Tropp (2006) obtained an effect size of +.21 between intergroup contact and reduced prejudice. Similarly, subjective well-being relates to both internal locus of control (+.25) and low neuroticism test scores (+.27, DeNeve & Copper, 1998). Likewise, in an analysis of more than 120 meta-analyses dealing with psychological assessment, the median effect size was +.27 (Meyer et al., 2001). Finally, the effect size of +.23 represents a Cohen’s d of +.47, which is recognized as a solid medium-sized effect in psychological research (Cohen, 1988).

**Improving RD predictability**

But clearly there is still room for improvement. First, the fit hypothesis should be furthered specified. For example, we determined fit by whether a particular behavior serves the group or the individual. But for these analyses, we had to assume crudely that individual acts of deviance served the individual and collective protest served the group. However, the decision to “tag” a local business with graffiti could represent a desire to express one’s individuality or a group norm (suggesting the proprietors are not welcome in the neighborhood). Similarly, the choice to attend a political protest could reflect one’s commitment to the collective goal or a
desire to join a friend at the rally. If future researchers measure respondents’ \textit{intention} to serve the group or their individual self, the fit between RD level and outcome could well be improved.

Second, very few affective RD measures in our data set include the discrete emotions of anger and resentment that RD theorists propose to be associated with RD (see Smith et al, 2008). The nine projects in which investigators actually asked respondents whether they felt angry or resentful about the disadvantaged comparison yield a strong average effect size of $+.34$ ($CI=+.26$ to $+.43$, $N=2,036$). Resentful anger is distinct from envy, jealousy or even more generic forms of anger (see Leach, Snider & Iyer, 2002, Smith et al, 2008). First, resentment should elicit a focus on the system that produces the inequity whereas envy should elicit a focus on the comparison targets and what they have. Second, resentment is often a publicly shared emotion that evokes notions of justice whereas envy is often a private and perhaps even shameful emotion (see Runciman, 1966). If future researchers measure resentful anger directly, RD predictability should improve.

Third, researchers should clarify why they expect RD to be related to a particular outcome. It makes sense that the angry resentment generated by an undeserved ingroup disadvantage would be directed toward outgroups. Anger is an “attack” emotion associated with behaviors directed toward others (Mackie, Devos & Smith, 2000). However, it is less clear how RD might be related to attitudes toward one’s ingroup. On the one hand, we could imagine that people who feel that they and those like themselves do not have what they deserve will identify more strongly with their group in contrast to other groups (perhaps leading to politicized identification; see Simon & Klandermans, 2001). On the other hand, people could feel that they do not have what they deserve \textit{because} of their group membership. This attribution would lead them to identify less strongly with their group (in preparation, perhaps, to leave the group for another group).
Similarly, it makes sense that anger and resentment would be related to individual “deviant” behavior, but less clear why it would motivate individual achievement (particularly if achievement is identified within the system as the most legitimate avenue of behavior). If a person feels angry and resentful about a disadvantage, it seems unlikely that they will be motivated to work hard within the same system that produced their undeserved disadvantage. As Fine and Rosenberg (1983) argue based on qualitative interviews with young adults who chose to leave high school, leaving school emerged as an active response to a system that treated them unfairly whereas staying appeared to be a more passive response to the same injustices.

Finally, the pathways between RD and physical and mental health outcomes (and associated attempts to “self-medicate” one’s discomfort with drugs and/or alcohol) is unclear. Should the anger and resentment associated with RD lead to depression, anxiety, heart disease, obesity or some combination? In part, this ambiguity reflects general questions about the relationship between specific emotional states and their psychological and physical consequences (Schnittker & McLeod, 2005). One difficulty for this area of research is the tendency to confuse anxiety about one’s status (e.g., concerns over how one compares with others, Schnittker & McLeod, 2005) and RD. In an organization or social system in which status hierarchies are unstable or uncertain, it is not surprising that higher status individuals report more stress related illness. However, such status anxiety should be distinguished from RD in which one’s (undeserved) disadvantaged position is a clear product of external circumstances. Indeed, if (individual) RD is related to physical and mental health, it should be associated with diseases linked to anger, not depression and anxiety. More importantly, it might not be a single RD experience that leads to poor physical and mental health, but cumulative experiences of RD and injustices that lead to poor outcomes (see Adler et al., 1994). In fact, health researchers describe a “weathering effect” in which the effects of social inequality on health increase as people get
older (Geronimus, Hicken & Keene, 2006). In other words, it might not be the experience of RD per se, but whether people’s attempts to address their disadvantages are successful that shape physical and mental health.
Limitations

Before considering the broader implications of this review, we should recognize four limitations to our conclusions. First, even after controlling for measurement quality, matched level of analysis and justice-related affect, the effect sizes remain remarkably heterogeneous. Although the heterogeneity supports our decision to use the more conservative estimates provided by the random effects model, it also suggests that the true effect of correctly measured RD is stronger than we are able to document.

The heterogeneity among effect sizes also indicates the importance for researchers to consider additional variables that might moderate or mediate the effects of RD on various outcomes. If researchers follow our suggestion to focus on measuring the discrete emotions associated with RD, we think it would be valuable to consider additional situational appraisals shown by emotion and RD theorists to be associated with anger and perceptions of deprivation (Smith et al., 2008). In particular, it is important to know who they believe is responsible for the deprivation, whether the process producing the deprivation was legitimate, whether the situation will change without any interference, and whether they feel powerful enough to confront the disadvantage.

A second limitation is our inability to infer causal relationships. Although our data set contains twelve longitudinal studies and three experiments, the great majority of these data come from cross-sectional surveys and questionnaires. There is no reliable difference between the average effect size for the 32 longitudinal tests and the 691 cross-sectional tests ($Q_{n} = +.06$, $p = .81$), but we do adopt the common theoretical assumption that feeling deprived leads to various reactions even though non-recursive causation is certainly possible. That is, one could argue that the various outcomes that we reviewed could lead to feeling deprived. For example, people’s awareness of their (collectively) undeserved disadvantage could follow as well as precede their
participation in collective protest (see Drury & Reicher, 2000). Or, perhaps, people who already feel depressed or anxious are more likely to notice and react to perceived undeserved disadvantages. In his ethnographic study of young people's criminal activity in a small town near London, Webber (2010) argues that small indiscretions committed by young people leads others to identify them as criminals, and this identity, in turn, prevents these young people from achieving their aspirations. In other words, it is not (economic) RD that leads to crime, but crimes that lead to RD. Ironically, Yang and his colleagues (2008) argue for the opposite relationship. They argue that greater exposure to material culture (measured as watching Western television shows) should increase RD (and perhaps lead to crimes of acquisition). In this review, we treated watching television as a form of escape from the RD experience in the same way that increased drug or alcohol use could be considered as forms of escape. We also treated crime, delinquency and counterproductive work behavior as reactions to RD. Again, our goal was to create the most inclusive dataset possible.

Still, we think there is a rich opportunity for experimental investigations of these relationships. For example, Callan, Ellard, Shead & Hodgins (2008) experimentally manipulated IRD by asking undergraduate students to calculate their monthly discretionary income in the context of comparison information from other undergraduates to investigate the effect of IRD on undergraduates’ intentions to gamble. And Smith and colleagues (1996) presented group members with distributions of participant payments that paid many more outgroup members in comparison to the ingroup as a manipulation of GRD.

Finally, our analysis suggests clear and continuing gaps within the RD literature. First, theoretically accurate multi-item measures are relatively rare, particularly for comparisons to one self or one’s group across time. Indeed, there is yet to be a proper test of Gurr’s (1970) original distinction between detrimental, aspirational and progressive forms of deprivation (in which
people’s ability to attain resources and the resources to which they feel entitled vary in different ways across time). Of course, one difficulty is the proper treatment of temporal comparisons. Are they a relevant and independent form of RD, or do they represent appraisals of situational stability (Mummendey et al., 1999) or feasibility (Crosby, 1976; Folger, 1987)? As part of their interviews in low income neighborhoods in Texas, Franzini & Fernandez-Esque (2006) asked respondents to compare themselves to other Mexicans and Anglos and also whether the chance that they or their children “can get ahead” was worse than average, average or above average. Such a combination of comparisons may offer the strongest and most accurate assessment of RD (see also Tyler, Boeckmann, Smith & Huo, 1997).

These data also suggest some unexpected but important patterns. First, comparisons to outgroup members strongly predicted internal states outcomes and the desire to escape or exit the situation. These comparisons more clearly represent (upward) contrasts between the self and target, as opposed to comparisons to other ingroup members that could be either assimilation or contrast comparisons (see Buunk, et al., 2001; Buunk et al., 2003). As Smith and Walker (2008) argue, comparisons to outgroup members could be an important developmental bridge between IRD and GRD – moving people from interpreting their disadvantage as a product of interpersonal circumstances to viewing their disadvantage as a product of intergroup relationships. Second, attitudes toward the system were as strongly related to IRD and temporal RD as to GRD. This pattern supports our suggestion that all forms of RD are associated with resentful anger and an interest in the external agent responsible for the deprivation, but it will be important for future research to validate this assumption.

**Future research**

How should we measure RD? Our meta-analytic results indicate an adequate measure of RD must include a clear comparison referent as part of the questions asked of respondents and
measure resentful anger. Researchers also must avoid the ambiguity implied by comparisons between individuals and potential outgroup members by determining whether respondents are thinking of themselves as group representatives or unique individuals. For example, Hafer and Olson (1992) asked working women the extent to which they felt resentful about women’s working situation compared to men’s working situation, and their answers strongly predicted their self-reports of political action ($r (69) = +.45$). Similarly, faculty members who reported feeling angry that faculty pay at their university compared to the pay of faculty members at comparable universities was worse than faculty members deserved reported more willingness to protest ($r (369) = +.33$, Smith et al, 2008). Both measures clearly capture intergroup comparisons.

In contrast, faculty members who reported feeling angry that their individual pay was worse than they deserved reported more stress ($r (369) = +.38$, Smith et al, 2008). Similarly, employees of two merged Korean telecommunications service companies who reported that compared to other employees they felt unfairly treated, worse off and more dissatisfied also reported more interest in leaving the company for another job ($r (274) = +.37$; Cho, 2003). In these examples, the measures clearly capture interpersonal comparisons.

It is tempting to propose that researchers simply measure injustice, deserving or anger directly without reference to particular comparisons. In a meta-analysis of 65 studies of collective action, the average effect size between collective action and non-comparative measures of injustice is $+.34$ (van Zomeren et al, 2008). But we think such an approach would lead researchers astray. What makes RD so useful is the recognition that those who should feel deprived by objective standards often do not feel deprived, and those who are not objectively deprived often feel that they are. It is the contextual and flexible nature of social comparisons that remind researchers that RD and injustice are not the property of a single person or group but rather the property of particular relationships. At the same time, people’s comparisons are not
completely divorced from reality (see Spears & Manstead, 1989). In fact, it is people’s “subjective expectations of objective probabilities” (Bourdieu, 1990, p. 59) that makes RD such a useful concept. When these subjective expectations shift and change due to imposed or chosen comparisons, we should predict increased or decreased action. To paraphrase Marx, it is only after people notice that their neighbors have flat screen televisions and new automobiles that they will feel deprived. Similarly, it is when students realize that they do not have access to the educational training that leads to better paying jobs that they will feel deprived. As previous researchers have described, collective and individual challenges to disadvantage often come from people who have more rather than fewer resources – recall Stouffer’s Army Airmen who revealed more RD while receiving far more promotions than the military police. If people do not experience RD either as individuals or group members, it seems highly unlikely that they would risk protest action (see Taylor and McKirnan’s (1984) five stage group model for a similar argument).

Our results call for personality and social psychologists to reconsider their relative neglect of RD. Measured properly, the variable is a significant predictor of a wide range of important outcome variables spanning collective action, individual deviance, and physical and mental health.
Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the authorship and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for this research and/or authorship of the article: RUI Grant 064500 from the National Science Foundation. We would like to thank all the researchers who generously shared their data with us.
Footnotes

Previous typologies often distinguish between normative (or system-facilitating) attitudes and behaviors and non-normative (or system-inhibiting) attitudes and behaviors. Although the distinction may seem obvious in theory, in practice such a distinction is more difficult to make. Is a faculty member’s choice to cancel office hours in response to a pay cut a normative or non-normative behavior? Is a protester’s choice to remain at a rally after being asked by police to leave a normative or non-normative behavior? Often, the assessment of a particular behavior as normative depends upon one’s group membership and place in the hierarchy. Although we later make choices that appear to support this distinction (e.g., the assumption that more RD should be associated with less support for the current political system, and the distinction between deviance and achievement behavior), we prefer to avoid labeling behaviors as violating or supporting particular norms.

2 Although we recognize that on occasion one can protest for individual reasons and steal for collective reasons.

3 We also explored a range of other possible predictors: publication year, respondent age, type of design, whether the RD or outcome measure had been used before, whether the sample represented a homogeneous group or not, whether the sample represented a traditionally disadvantaged group or not, whether the RD dimension represented economically-based differences or not and whether the outcome was self-reported or not. Details about these analyses are available from the authors upon request.
References


Structural validity and generalizability of a referent cognitions model of turnover intentions.


*Bourguignon, D., Seron, E., Yzerbyt, V., & Herman, G. (2006). Perceived group and personal


social comparison while learning social skills in groups. *Group Dynamics: Theory, Research and Practice, 11*, 140-152.


*Ghosh, E. S. K., Kumar, R., & Tripathi, R. C. (1992). The communal cauldron: Relations between Hindus and Muslims in India and their reactions to norm violations. In R. deRidder & R. C. Tripathi (Eds.), *Norm Violations and Intergroup Relations* (pp. 70-89). Oxford:


*Kim, J., Sorhaindo, B., & German, T. E. (2006).* Relationship between financial stress and workplace absenteeism of credit counseling clients. *Journal of Family Economic Issues, 27,*


*Moore, D. (2008). Willingness to participate in social action: The beliefs that sustain it and the factors that shape it. Unpublished manuscript, Department of Sociology and Anthropology, Hebrew University of Jerusalem, Israel.


9, 111-117.


*Schruijer, S. G. L. (1992). On what happens when Dutchmen and Turks violate each other's


*Traupmann, J., Hatfield, E., & Sprecher, S. (1981). *The importance of fairness for the marital*
satisfaction of older women. Unpublished manuscript, Department of Psychology, University of Hawaii at Manoa.


...
relative deprivation and militancy.] Unpublished manuscript, Department of Sociology, University of South Africa, Pretoria, South Africa.


*Wong, N. K., & Walker, I. (1994). The relevance of the distinction between egoistic relative deprivation and fraternalistic relative deprivation in predicting environmental activism.


*Zoogah, D. B. (2010). Why should I be left behind? Employees’ perceived relative deprivation
Figure 1. Funnel Plot of Sample Size by Fisher’s Z
Figure 2. Mean Effects of Various Test Subsets by Percentage of Variance Accounted For
Table 1. Summary of effect sizes for RD measures and outcomes.

<table>
<thead>
<tr>
<th>Sample</th>
<th>r</th>
<th>95% CI</th>
<th>Z</th>
<th>k</th>
<th>N</th>
<th>Qw</th>
<th>Tau</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent studies</td>
<td>.106</td>
<td>.084-127</td>
<td>9.69</td>
<td>210</td>
<td>143188</td>
<td>3018.60</td>
<td>.142</td>
</tr>
<tr>
<td>Independent samples</td>
<td>.144</td>
<td>.128-.161</td>
<td>17.00</td>
<td>293</td>
<td>186073</td>
<td>3264.47</td>
<td>.128</td>
</tr>
<tr>
<td>Separate RD measures</td>
<td>.134</td>
<td>.121-.148</td>
<td>19.35</td>
<td>421</td>
<td>243733</td>
<td>4090.63</td>
<td>.124</td>
</tr>
<tr>
<td>Internal states</td>
<td>.173</td>
<td>.152-.193</td>
<td>16.19</td>
<td>188</td>
<td>135198</td>
<td>2393.82</td>
<td>.129</td>
</tr>
<tr>
<td>Individual behavior</td>
<td>.118</td>
<td>.097-.140</td>
<td>10.68</td>
<td>126</td>
<td>81474</td>
<td>988.20</td>
<td>.106</td>
</tr>
<tr>
<td>Intergroup attitudes</td>
<td>.115</td>
<td>.097-.134</td>
<td>12.00</td>
<td>299</td>
<td>152366</td>
<td>3619.76</td>
<td>.149</td>
</tr>
<tr>
<td>Collective behavior</td>
<td>.148</td>
<td>.115-.181</td>
<td>8.66</td>
<td>99</td>
<td>49242</td>
<td>1268.18</td>
<td>.157</td>
</tr>
</tbody>
</table>

Note. The mean effects and confidence limits listed in this table have been transformed back to the r metric from the z transformed estimates obtained in the original analyses.
Table 2. Comparison of cognitive and affective RD measures.

<table>
<thead>
<tr>
<th>Sample</th>
<th>r</th>
<th>95% CI</th>
<th>Z</th>
<th>k</th>
<th>Q_w</th>
<th>Tau</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive RD</td>
<td>.077</td>
<td>.053-.100</td>
<td>6.34</td>
<td>124</td>
<td>1494.95</td>
<td>.117</td>
</tr>
<tr>
<td>Affective RD</td>
<td>.174</td>
<td>.128-.220</td>
<td>7.33</td>
<td>60</td>
<td>679.62</td>
<td>.169</td>
</tr>
<tr>
<td><strong>Between-Classes effect</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Independent samples</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive RD</td>
<td>.129</td>
<td>.110-.145</td>
<td>7.94</td>
<td>173</td>
<td>1465.87</td>
<td>.100</td>
</tr>
<tr>
<td>Affective RD</td>
<td>.186</td>
<td>.141-.231</td>
<td>14.17</td>
<td>87</td>
<td>1469.21</td>
<td>.206</td>
</tr>
<tr>
<td><strong>Between-Classes effect</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Separate RD measures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive RD</td>
<td>.116</td>
<td>.102-.130</td>
<td>15.28</td>
<td>264</td>
<td>1962.59</td>
<td>.096</td>
</tr>
<tr>
<td>Affective RD</td>
<td>.165</td>
<td>.134-.195</td>
<td>10.38</td>
<td>157</td>
<td>1931.59</td>
<td>.184</td>
</tr>
<tr>
<td><strong>Between-Classes effect</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. The mean effects and confidence limits listed in this table have been transformed back to the r metric from the z transformed estimates obtained in the original analyses.
**Table 3.** Comparison of “matched” and “mismatched” RD and outcome measures.

<table>
<thead>
<tr>
<th>Sample</th>
<th>r</th>
<th>95% CI</th>
<th>Z</th>
<th>k</th>
<th>Q_w</th>
<th>Tau</th>
</tr>
</thead>
<tbody>
<tr>
<td>All RD comparisons</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Matched levels of analysis</td>
<td>.166</td>
<td>.151-.182</td>
<td>20.80</td>
<td>333</td>
<td>4,175.79</td>
<td>.130</td>
</tr>
<tr>
<td>Mismatched levels of analysis</td>
<td>.113</td>
<td>.089-.137</td>
<td>9.18</td>
<td>167</td>
<td>2,001.90</td>
<td>.142</td>
</tr>
<tr>
<td><strong>Between-Classes effect</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affective RD measures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Matched levels of analysis</td>
<td>.201</td>
<td>.167-.235</td>
<td>11.17</td>
<td>119</td>
<td>1,683.47</td>
<td>.183</td>
</tr>
<tr>
<td>Mismatched levels of analysis</td>
<td>.123</td>
<td>.076-.169</td>
<td>5.10</td>
<td>77</td>
<td>1,288.43</td>
<td>.200</td>
</tr>
<tr>
<td><strong>Between-Classes effect</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affective RD measures with quality controls</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Matched levels of analysis</td>
<td>.230</td>
<td>.174-.285</td>
<td>7.89</td>
<td>43</td>
<td>1098.35</td>
<td>.186</td>
</tr>
<tr>
<td>Mismatched levels of analysis</td>
<td>.111</td>
<td>.011-.210</td>
<td>2.164</td>
<td>27</td>
<td>1080.65</td>
<td>.264</td>
</tr>
</tbody>
</table>
| **Note.** The mean effects and confidence limits listed in this table have been transformed back to the r metric from the z transformed estimates obtained in the original analyses. Comparisons to outgroup members and attitudes toward the ingroup are not included. Quality controls exclude outcomes measured as difference scores, single item RD measures and convenience samples.