Social Judgment and Decision Making

10

More Than an Artifact
Regression as a Theoretical Construct

KLAUS FIEDLER AND JOACHIM I. KRUEGER

> Regression to the mean is like the weather: Everybody talks about it, but few of us do anything about it. However, it is unlike the weather, because most of us fail to recognize it, even when it hits us in the nose.
>
> **Lewis R. Goldberg (1991, p. 181)**

INTRODUCTION

One scientist's research finding is often another's artifact. Upon reflection, a presumed artifact can turn out to be of great scientific value. We explore this possibility with regard to phenomena of statistical regression. Regression effects pervade social science and especially the psychology of judgment and decision making. Yet, the typical response is to see these effects as threats to a study's internal validity (Campbell & Kenny, 1999) or to ignore them altogether. In contrast, we will show that regression can be a powerful explanatory construct with important theoretical implications.

Regression and the Phenomenology of Unusual Outcomes

Scientific Discoveries Consider the common and rather distressing finding that empirical effects tend to weaken with replication (Jennions & Møller, 2002). When initial discoveries fail to replicate with their original strength, it is tempting to search for causal explanations, such as differences in the study design, the location of study, or the type of participant sample. Such a precommitment to a causal perspective hinders a proper appreciation of the statistical law of regression.

Every empirical finding is characterized by a nonzero effect size and imperfect measurement reliability. Assuming a retest reliability of $r = .67$, for example, the retest scores x' will be regressive by one third ($1 - r = .33$) relative to the original scores x. An individual i whose initial z-standardized score is $z(x_i) = 3$, that is, whose score lies three standard deviations above the mean, will probably have a retest score of $z(x'_i) = 2$. Likewise, at the study level, an initial mean difference between two study groups of, say, $d' = z_{\text{Experts}} - z_{\text{Laypeople}} = 1.00$ will shrink by the same scaling factor to $d'' = .67$. Although replication studies may sometimes yield stronger results than the original studies because of improvements in the precision of measurement, such an outcome is unlikely. The original findings most likely to stimulate replication studies are those that have large effect sizes. This is so because large effects are most likely to be published, as a premise for replication (Fiedler, 2011). Yet, the effect sizes of the replications tend to be smaller than the original effects, reflecting both the regressive shrinkage of scores at the individual level and the regressive decline of typically inflated (rather than underestimated) initial effects at the study level.

Bidirectional Regression Why does unreliability not lead as often to increased effect sizes? This question reflects a basic psychological barrier to understanding the regression principle. Even Galton (1886) first misunderstood his data on heredity (Stigler, 1999). Galton found that very tall fathers are taller and that very short fathers are shorter than their sons. This pattern looked as though the height of men becomes more uniform over generations. But Galton also noticed that very tall sons are taller and that very short sons are shorter than their fathers. Now it seemed as though the height of men becomes less uniform. Both trends cannot be true. The logic of regression explains that it is only the variation of the prediction estimates that shrinks, whereas the predictor levels are fixed at their original values. Given two random variables X and Y with equal variance and a less than perfect intercorrelation ($|r_{XY}| < 1$), the variance of estimates of X shrinks when plotted against fixed levels of Y, but the variance of estimates of Y also shrinks when plotted against fixed levels of X (Furby, 1973).

Now suppose all studies were published and followed up with attempts to replicate. Replications would yield less extreme results when plotted against original results. Yet, the original results would appear less extreme if plotted against replications. The regression law does not imply a real difference between originals and replications. Replications only appear to yield smaller effects because they are typically plotted against (i.e., conditionalized on) the original results. Analogous to the Galton story, one might produce the opposite phenomenon in reverse analyses. Plotting Y against X means to take observed X values for granted, despite the error with which X is measured. Given that $|r_{XY}| < 1$, any value on the X axis belongs to a true value that is less extreme. Consequently, the Y estimates that correspond to fixed levels of X must be less extreme. This does not reflect any essential difference in the variability of X and Y but merely the fact that X is fixed and only Y is allowed to shrink with decreasing r_{XY}.

Everyday Success and Failure The regression-replication trap has parallels in everyday judgment and decision making. When a political party or candidate wins an election by a landslide, the rational voter expects a more modest outcome the next time around. Landslides are by definition extreme and unusual. Politicians, pundits, and casual observers who are blind to regression will rush to explain both the landslide and the close election following it in causal terms. If they appreciated regression, they would estimate the degree to which election outcomes lack perfect reliability (e.g., by considering the historical record), make a regressive forecast for the next election, and then compare the observed result with the predicted one.

Similar scenarios arise in business, sports, and popular culture. Often, people are hired, promoted, or married at a time when they are at their personal, but unreliable, best. The regression fallacy is to think that this stellar moment reflects a new high plateau ("He/she will never change!") or worse yet, the first step in a journey to ever-improving excellence. Such hopes may be fulfilled in a few cases, but these are hard to identify *before* the fact. The typical scenario is that of regression. The moment of promotion, winning a contest, marriage, and so on, is one in which performance or fortune (i.e., the "evidence") surpasses a high threshold. It is difficult to see that this evidence is but a sample drawn from a latent population. It is the central tendency and the variability of that latent population that characterize the person, not the extraordinary sample retained *because* it passed the threshold.

Insensitivity to regression and the need to explain random variation breeds disappointment, damages careers, and breaks families. Take the plight of well-intentioned, but regression-blind, educators (Stelzl, 1982). Novice teachers often assume that students' academic achievement and social behavior responds to positive reinforcement, and that punishment of failure and misbehavior is ineffectual and undesirable. Failing to take stochastic factors into account, educators who most enthusiastically reward the highest student achievement and the most positive conduct will notice that subsequent performance and conduct are often less outstanding. They may then falsely conclude that praise damages performance. If teachers then resort to punishing poor performance and behavior, they will notice improvements. Over time, regression blindness turns optimistic reinforcers into cynical punishers. This deplorable sequence of events would not occur if teachers could estimate, however crudely, the random element in student behavior, adjust for it, and evaluate the effectiveness of their feedback by comparing student behavior with informed predictions that take regression into account.

How Much Regression? We have emphasized the need to draw rational inferences by comparing observations with predictions informed by the regression logic, not with selective or haphazard past observations. The direction and amount of change that can be expected from regression can be quantified precisely, assuming a standard normal distribution. Let $Y = (X - \mu)$ be the deviation of an observed value X from the population mean μ. The given deviation score of Y will regress to $Y^\circ = Y \times r$, whereby r is the reliability of measurement. The strength of the regression shift $Y^\circ - Y$ is $-Y \times (1 - r)$. Regression is an inverse joint function of the extremity of an original, error-prone measurement, multiplied by its unreliability $(1 - r)$. It is negative when original measures are above the mean (i.e., when $Y > 0$) and positive when original measures are below the mean. It increases linearly with both extremity ($|Y|$) and unreliability $(1 - r)$.

These implications are relevant for the theoretical analysis of many putative illusions and biases in judgment and decision-making. Even when the parametric regression model is not applicable because the assumption of a metric normal distribution is not met, or because prior knowledge or extraneous influences cause regression to move to some other value than the arithmetic mean, the qualitative and ordinal implications remain: Regression counteracts the apparent trend observed in any (error-prone) deviation measure, and its effect strength increases with extremity and unreliability.

REGRESSION AS A KEY CONCEPT IN PSYCHOLOGICAL EXPLANATIONS

As noted earlier, regression effects are often dismissed as artifacts. Psychologists and other scientists have been slow to reflect on the theoretical potential of regression (cf. Rulon, 1941). One reason for this neglect may lie in implicit philosophical assumptions about what qualifies as an adequate explanation of human judgment and behavior. Most social and personality psychologists seek to explain behavior in terms of internal (intrapsychic) factors such as intentions, needs, or traits, or external factors such as salient stimuli, social influence, or social roles and norms (Ross & Nisbett, 1991). They rarely consider unsystematic error as a source of systematic effects. By contrast, we argue that acceptance of regression as an explanatory construct can enrich psychological theories.

We elaborate this idea in the context of prominent research findings, devoting four subsections to different ways in which unsystematic regression error can generate systematic biases. These four variants of theoretical regression effects, which lie at the heart of different classes of theoretical explanations, are regression in longitudinal assessment, regression in conditional reasoning, differential regression, and regression approaches to subadditivity.

Regression in Longitudinal Assessment

Timing of Interventions Regression effects pervade longitudinal research in developmental psychology, education, and quality control studies for intervention programs (E. Thorndike, 1924). Intervention studies in business, education, or in the field of public health (e.g., rewards, punishments, therapy, compensatory training, etc.) rarely use designs with random assignment. Their timing depends on abnormal or extreme states or events. The abnormal moments, which elicit interventions (or more generally, actions and reactions), entail regression, and hence any apparent impact of an intervention is ambiguous. The evident success of an intervention may in part or fully be the result of regression.

Eysenck (1965) made this point in his early critique of research on the efficacy of psychotherapy. Even when most patients improve with time, there is no reason for celebration if a similar rate of improvement may be observed as spontaneous remissions in a nonintervention control group. Although psychotherapy research has become more refined since Eysenck's days (Weisz, McCarty, & Valeri, 2006), such control conditions remain rare. Campbell (1996) observed that therapy

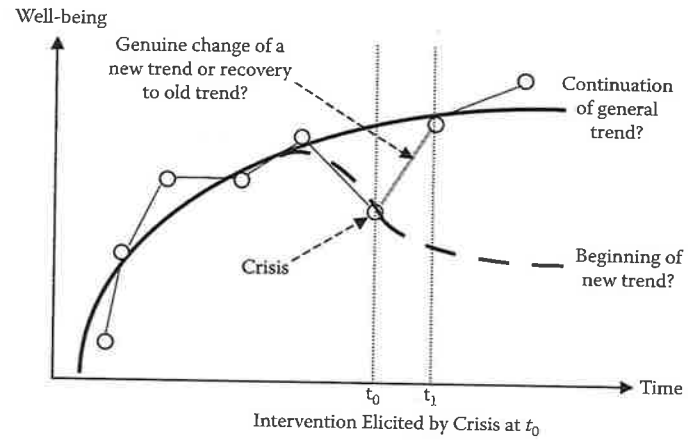

Figure 10.1 Regression provides an alternative account of improvement after an intervention elicited by an unusually low measurement point, or crisis at t_0. The increase between t_0 and t_1 need not reflect the causal impact of the intervention intended to counteract a new downward trend (dashed line). It may simply reflect a normal return to the original trend (solid line), which may be causally independent of the intervention.

is typically introduced or changed at a point of crisis or abnormal peak in the patients' symptoms. If the history of the symptoms prior to intervention is ignored, it is impossible to distinguish the causal impact of the intervention (dashed trend in Figure 10.1) from a regressive return to the old long-term baseline (gray line leading back to the solid curve).

Misattribution and Superstition Practical intelligence involves the ability to cope with "statistical deceptions." Campbell (1996) considered the case of "Compound Q," which was promoted as a drug to manage AIDS. Although the necessary data were available, researchers did not analyze an extended time series to determine whether the success of the intervention was due to a genuine improvement over the existing trend or merely to a return to the normal trend from an abnormal crisis that motivated the intervention. Campbell cautioned that blindness to regression can be motivated. Agents who are accountable for expensive interventions or new medical developments have little interest in designing studies sensitive enough to test alternative explanations.

Much like scientists' misinterpretations of intervention effects may reflect both the regression trap and vested interests, everyday attribution errors are also subject to both influences. Voters have to judge whether a government's policies have been successful. They must distinguish between illusory failure due to regression (when a government started at a time of abnormal prosperity that would regress statistically to a more modest level) and genuine failure (a reduction in welfare beyond regression). Facing this ambiguity, voters' attributions typically stay consistent with their preferred political standpoint.

People often try to control (and improve) the outcomes in their lives. Some behavior, such as intercessional prayer or taking substances without demonstrated medicinal value, is futile considering the best evidence; yet regression (from a crisis that causes the behavior) makes it look effective, and the person choosing these interventions is vulnerable to an illusion of control (Langer, 1975). The sequence of poor condition, intention to act, taking action, and experiencing improvement, makes it virtually irresistible to see oneself as the author of improvement.

People are not blind to variations in the control they have over their outcomes (Thompson, Armstrong, & Thomas, 1998). With the correlation between true and perceived control less than perfect, regression guarantees that the overestimation of control will be at a maximum when actual control is minimal. The only way to err is to overestimate control. Conversely, if a study was designed in which actual control was perfect, the only way to err would be to underestimate control.

These examples illustrate the manifold ways in which regression effects can account for illusory or superstitious attributions. A final illustration for the counterintuitive nature of regression can be found in the stock market, which is essentially regressive. By analyzing the values of a set of different shares over 15 years, Hubert (1999) showed that investors would have been better off always buying the shares with the lowest annual average profit than buying always the shares with the highest annual profit. Accordingly, successful investment means to anticipate reversals in trends, buying just before an upturn and selling just before a downturn.[1] Because, however, stock prices are not sampled from a population with stable parameters, the estimation of regression to a moving mean becomes an intricate problem.

Moral and Political Reasoning Blindness to regression lies at the heart of many moral, ethical, and social-political judgments. Consider again a prominent research example. To study the effectiveness of educational or economical programs intended to compensate for existing inequalities in social-economic status or between gender groups, researchers often compare matched groups with "equivalent" test scores, one from an advantaged group and one from a disadvantaged group (McNemar, 1940; R. Thorndike, 1942).

If posttraining performance is lower in the disadvantaged group (e.g., immigrant, minorities, working class people), this is often attributed to the ineffectiveness of the program, or the "genetic" inferiority of the disadvantaged group. Such inferences can be groundless because regression plays a subtle yet powerful role. Note that the two groups are matched only at the sample level. Matching samples from two different populations means that the sample of the advantaged group is likely to underestimate the real performance, whereas the sample drawn from the disadvantaged group is likely to overestimate the real performance. Independent of any causal influence of the training program, a shift toward their respective population means can be expected. This shift needs to be statistically controlled before social interventions can be evaluated. Careful analysis might show that a program has worked, although the apparent sample difference appears to refute this claim.

Regression in Conditional Reasoning

The defining characteristic of regression is the lack of a perfect correlation (Campbell & Kenny, 1999). Temporal instability is but one source of regression effects. In concurrent measurement, the lack of a perfect correlation can signal a lack of reliability, validity, or accuracy. The psychology of judgment and decision making is replete with findings computed as differences between observed and predicted values. These differences are typically chalked up to the "limited capacity" of the human mind to make rational decisions. Upon reflection, however, many of these differences can be understood as a regressive transformation of random variation into systematic discrepancies, or "effects."

Overconfidence In a typical study, participants receive a list of knowledge questions (e.g., Which river is longer, the Mississippi or the Nile?). For each, they select an answer and indicate their degree of confidence. Judgments are considered well calibrated if the degree of confidence matches the probability of being correct. Overconfidence means that people overestimate the probability of being correct.

Of course, subjective confidence levels vary. Likewise, there is variation in item difficulty. By definition, questions that few (many) people answer correctly are difficult and questions that many people answer correctly are easy. Confidence and accuracy are positively correlated as people have some insight into the quality of their judgments. Yet, when accuracy is plotted against confidence, very high confidence is usually too high, and very low confidence is usually too low (see Figure 10.2). Therefore, subjectively easy items (with high or extremely high confidence) entail overconfidence, whereas subjectively difficult tasks (with low and extremely low confidence) entail underconfidence (Juslin, Winman, & Olsson, 2000).

Studying overconfidence requires a decision as to which criterion will be regressed on what predictor. When regression is done both ways, the same data can show both overconfidence and underconfidence (Dawes & Mulford, 1996; Erev, Wallsten, & Budescu, 1994; Soll, 1996). Consider easy tasks producing high confidence ratings. When the accuracy on these tasks is plotted against confidence, the accuracy corresponding to the highest confidence levels will be less extreme, thus exhibiting overconfidence. When, however, confidence is plotted against accuracy, the confidence for the easiest or most accurate items will be less extreme, thus showing underconfidence. Conversely, the most difficult items with the lowest accuracy will produce overconfidence when confidence is plotted against given accuracy levels. The same items produce underconfidence when the accuracy corresponding to extremely low confidence ratings is actually less extremely low.

Optimistic Bias Research on self-enhancement, or the so-called illusion of invulnerability, parallels some of the issues illustrated with confidence research (Alicke & Sedikides, 2009). In this paradigm, self-judgments are typically plotted against objective measures, and it is noted that most self-assessments are more favorable than actual performance. Like the hard–easy effect in overconfidence, this effect is most pronounced for low performers. In contrast, high performers underestimate themselves, though less extremely so (Krueger & Mueller, 2002). To

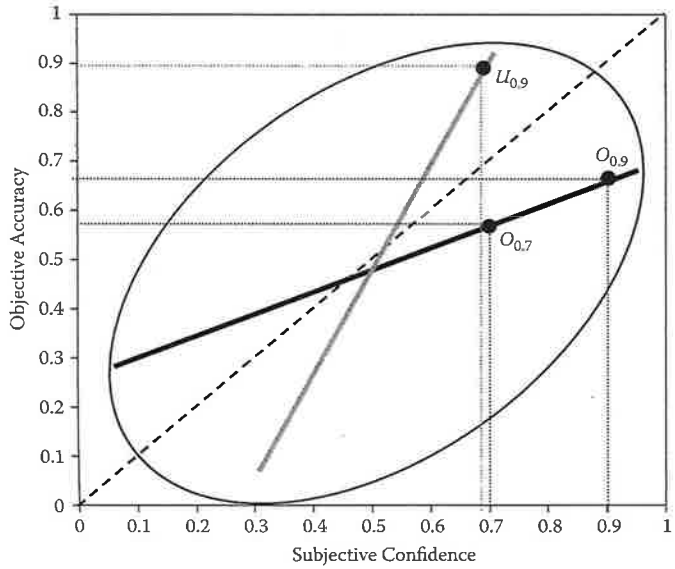

Figure 10.2 The bidirectional nature of regression creates overconfidence and underconfidence in the same data. When objective accuracy is plotted against given values of subjective confidence, overconfidence is apparent in the fact that the objective analogues of high subjective values are clearly lower (see data points $O_{.9}$ and $O_{.7}$). In contrast, underconfidence is apparent when subjective confidence estimates are plotted against given objective accuracies. The subjective analogue of high accuracy (see $U_{.9}$) is clearly lower.

account for this regressive pattern, there is no need to attribute different mental capacities or motives to high versus low performers, as Kruger and Dunning (1999) did. The logic of regression alone predicts these opposite judgment biases for high and low performers (Krueger & Wright, 2011). The same logic also predicts that the optimistic bias of low performers can be turned into apparent pessimism if the data are analyzed the other way around. On tasks with extremely low self-attributed percentile values, the actual percentile can be expected to be relatively higher.

A mean-level optimism effect can also be explained with the assumption that judgments under uncertainty do not regress to the midpoint of the response scale but to some optimistic default value above the midpoint of the scale. Although such a response tendency might be called an optimistic bias, it need not be explained in terms of self-serving motives. The same optimism is extended to friends (Epley & Whitchurch, 2008; Krueger, 1998) or other members of the social group (Klar & Giladi, 1997).

Differential Regression as a Source of Biased Comparisons

We have reviewed judgmental biases that reflect a failure to take the conditional nature of regression into account. Depending on whether we observe *Y* conditional on *X* or *X* conditional on *Y*, regression will render *Y* or *X*, respectively, less variable and less extreme. According to what might be called *Goldberg's Rule*, "The variance of our predictions should never be larger than that of the criterion we seek to predict. (Never, not just hardly ever)" (Goldberg, 1991, p. 181). Considering the correlation in only one direction may lead to premature conclusions about cognitive or motivational biases.

We now turn to phenomena that arise when two or more sets of judgments are regressive to different degrees. The result is that equivalent targets are judged differently, which is traditionally attributed to perceiver's biases in attention or motivation. Differential regression offers a more parsimonious account. The central idea is that judgments are less extreme for those targets for which the available observations are less reliable. As reliability increases with the size of the sample, one may expect that judgments based on many observations will be more extreme than judgments based on few observations.

Imagine a person playing the role of a classroom teacher (Fiedler, Freytag, & Unkelbach, 2007; Fiedler, Walther, Freytag, & Plessner, 2002). Sixteen simulated boys and girls are presented on the computer screen. Across several sessions, the teacher is giving lessons on different subjects. For each lesson, there is a list of questions on a pull-down menu. On every trial, the teacher selects a question for the class. Some students raise their hands, the teacher calls on one, and the answer is either right or wrong. The individual students differ in their percentage of correct responses (the ability parameter) and their percentage of raising their hand (the motivation parameter). Teachers are asked to assess performance over many trials and to grade all students on both ability and motivation at the end of each lesson.

Figure 10.3 shows the average estimates of student ability as a function of their actual ability parameter (20% versus 50% versus 80%) and successive lessons (from 1st to 4th). Before the first lesson, in the absence of any observations, a reasonable teacher assigns the same intermediate scores to all students. This state of indeterminacy reflects maximal regression, as indicated by the dashed horizontal regression line. After the first session, the learning of student-specific performance remains incomplete, but the slope is already visible. Nevertheless, real parameters of 20% and 80% regress to middling estimates of roughly 46% and 55%, respectively. As the assessment process becomes more complete, the degree of regression declines, reflecting the increasing extraction of judgment-relevant data. Yet, even after many lessons regression is not eliminated. Teachers' judgments never correlate perfectly with student parameters, due to uncontrollable sources of unreliability such as forgetting, attention loss, or fatigue.

Now consider two students that differ in the degree of regression because they are afforded unequal opportunities to learn about their performance. Students with the same high ability parameter (80%) could differ in how often they raise their hand. Assuming unbiased teacher attention, this results in a larger sample of observations or "learning trials" for the more motivated student. However, as the amount of learning determines reliability and regression, this means that high ability will be more apparent for the more motivated student simply because of differential regression. No motivational or cognitive anomalies need to be postulated to predict and explain this basic judgment bias.

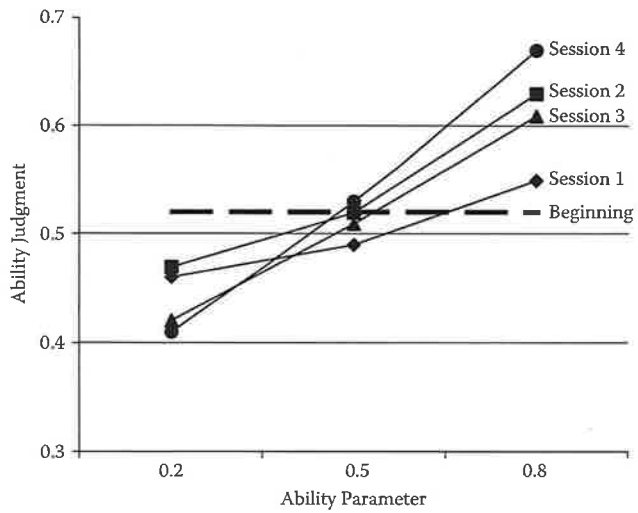

Figure 10.3 Regressive underestimation of high-ability students (0.80) and overestimation of low-ability students (0.20) in simulated classroom studies by Fiedler et al. (2002). The degree of regression decreases with learning experience, from session 1 to session 4.

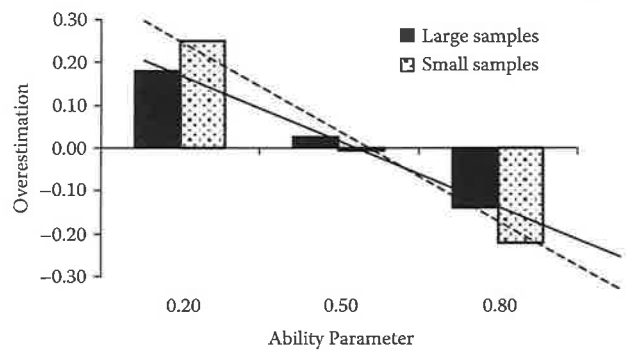

Figure 10.4 Regressive underestimation of high-ability students (0.80) and overestimation of low-ability students (0.20) in simulated classroom studies by Fiedler et al. (2002). Judgments based on small samples are more regressive than judgments based on large samples.

The results plotted in Figure 10.4 show that the same level of student ability led to more extreme judgments when the sample of observations was large rather than small. Small samples created a disadvantage for smart students, whose high ability was underestimated due to enhanced regression, whereas poor students profited from enhanced regression, which prevented teachers from fully recognizing their low ability (Fiedler et al., 2002).

Confirmation Bias and Positive Testing Differential regression can also explain findings traditionally seen as instances of confirmation bias (Nickerson, 1998; Snyder, 1984). In the simulated-classroom study, teachers were asked to test the hypothesis that boys are superior in math and physics, whereas girls are superior in language. As expected, these instructions triggered a positive-testing strategy (Klayman & Ha, 1987). Focusing on the students representing the focal hypothesis, most teachers gathered more science-related information from boys and more language-related information from girls. As a consequence, judgments of boys and girls were unequally regressive, leading to less positive evaluations of smart girls than equally smart boys in science, and to less positive evaluations of smart boys than equally smart girls in language. These differences in evaluation did not emerge among teachers who gathered samples of the same sizes. In other words, differences in evaluation depended on the regressiveness of available information, not on biased information processing.

Confirmation biases are often attributed to the undesirable effects of stereotypes. Hence, the case for the differential regression hypothesis is particularly strong if its signature pattern of results can be obtained even when it is pitted against the stereotype effect. Indeed, when teachers draw larger samples of girls in science and boys in language, the resulting differences in the evaluation of equally smart boys and girls run counter to the stereotype (Fiedler et al., 2002; Fiedler, Walther, & Nickel, 1999).[2]

Comparing the Self to the Group Average Differential regression also casts a new light on optimistic biases. Given the general response bias toward high performance estimates and given that people know more about themselves than about others, it follows that judgments of the self are less regressive—and thus more positive—than judgments of others. The regression account correctly predicts a reduced better-than-average effect when the amount of information about others increases (Moore & Small, 2007). By the same logic, a worse-than-average effect can result due to greater self-knowledge, when people compare themselves to the average other with regard to negative attributes (Klar & Giladi, 1997).

Illusory Devaluation of Minorities The devaluation of minority groups is a robust social-psychological finding that has generated many explanations (Dovidio & Gaertner, 2010). In Hamilton and Gifford's (1976) influential work on illusory correlations, participants receive behavior descriptions associated with members of two groups, denoted A and B. For Group A, the sample of behaviors is twice as large as for Group B. As positive behaviors are twice as frequent as negative behaviors in both groups, the correlation between group and valence is zero. Nevertheless, the prevalent positivity of behavior is more readily apparent for the majority than for the minority, as evident from frequency estimates, group impression ratings, and cued recall of behavior-group associations in countless experiments (Fiedler, 2000; Mullen & Johnson, 1990).

Conventional explanations of illusory correlations focus on cognitive or motivational biases, such as selective memory for the most infrequent combination (i.e., negative behavior by the minority), or a motive to give meaning and identity to

groups (McGarty, Haslam, Turner, & Oakes, 1993). However, differential regression is sufficient to produce the effect (Fiedler, 1991). Given more learning trials for the majority than the minority, it is no wonder that a generally high rate of positive behavior is more readily extracted for the former than the latter group (Fiedler, 1996; Klauer & Meiser, 2000).

Enhanced Regression Effects Through Subadditivity

A particularly intriguing class of biases in judgment and decision making arises when superordinate categories are split into subcategories or, conversely, when subcategories merge into superordinate categories. In this context, we encounter subadditivity. The overall subjective quantity $s(A + B)$ that results from merging two subsets or subcategories A and B is smaller than the sum of the subjective component quantities $s(A) + s(B)$. Subadditivity holds for both utilities (Morewedge, Gilbert, Keyzar, Berkovits, & Wilson, 2007) and probabilities (Fiedler & Armbruster, 1994; Tversky & Koehler, 1994). Gaining two times $1000 is worth more, subjectively, than gaining $2000. Tolerating an unpleasant physical examination for 10 minutes is less aversive than tolerating two 5-minute unpleasant examinations on separate days. The estimated risk or likelihood of dying from an accident is smaller than the summed estimate of the likelihood of dying in a traffic accident or a sports accident or a household accident, or a work accident, or a disjunctive list of other accidents.

Subadditivity entails so-called unpacking effects (Rottenstreich & Tversky, 1997; Tversky & Koehler, 1994) or category-split effects (Fiedler, 2002; Fiedler & Armbruster, 1994). The subjective likelihood or utility of a superordinate category (e.g., the number of Japanese cars on U.S. streets) can be increased by unpacking the category into two or more subcategories (e.g., number of Mazda, Toyota, Mitsubishi, Nissan, Honda, or Suzuki cars on U.S. streets). Conversely, packing several subcategories into a superordinate category reduces the overall subjective quantity.

Subadditivity can be derived from the negatively accelerated shape of most psychophysical functions. Large quantities are disproportionally underestimated because the subjective analogues of objective quantities (such as lightness, loudness, or monetary gains) do not increase linearly. The example in Figure 10.5 shows that adding $a + b$ yields a high quantity on the abscissa, the relative underestimation of which is much greater than for the two smaller component quantities together, as evident from the deviation between the subjective function and the dashed linear diagonal.

The Weber–Fechner law captures this psychophysical phenomenon. The difference threshold Δx (i.e., the increment in the stimulus quantity required to notice an increment) increases linearly with the absolute level x of the quantity. With the ratio $\Delta x/x$ constant, an increasingly larger objective input is needed to produce an identical subjective output increment Δx (i.e., a just noticeable difference).

A similar rule can explain why decelerating functions, implying subadditivity, are common in many other areas beyond psychophysics. In prospect theory (Kahneman & Tversky, 1979), the amount of objective gains (losses) needed to add a constant increment in subjective value increases with the absolute level of gains (losses). In

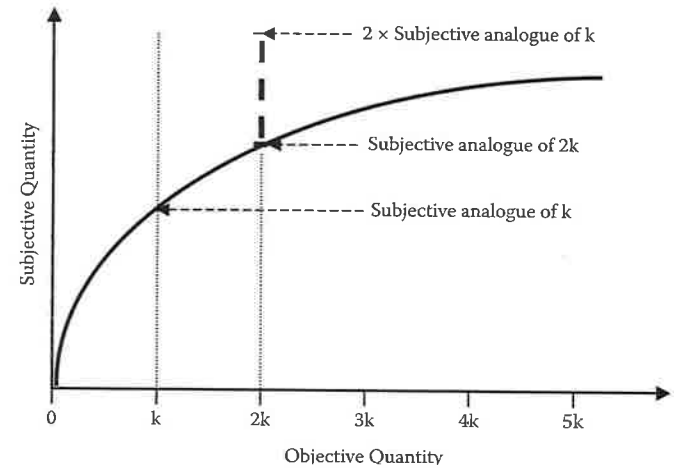

Figure 10.5 Graphical illustration of subadditivity created by negatively accelerated functions: Twice the subjective analogue of k is clearly more than the subjective analogue of 2k.

psychological statistics, too, the increment in sample size required to increase the resulting reliability by a constant factor increases with absolute sample size.[3]

As a rule, the dynamically changing discrepancies between subjective and objective quantities—in psychophysics, utility and subjective probability assessment—entail regression effects, that is, overestimation (underestimation) of small (large) quantities that results from the splitting (merging) of stimulus categories. In a recent study by Fiedler, Unkelbach, and Freytag (2009), participants estimated the number of butterflies of different types they had seen in an extended stimulus series. In one experiment, the objective presentation frequencies for butterfly types A, B, C, and D were 4, 10, 16, and 22, respectively. Frequency estimates were regressive because subjective and objective frequencies were not perfectly correlated. Replicating a standard result in memory for frequencies (cf. Fiedler, 2002; Sedlmeier, 1999), participants overestimated the frequencies of small categories (A and B), and underestimated the frequencies of large categories (C and D). Over- or underestimation was stronger for extreme categories (A and D) than for moderate categories (B and C). When the other determinant of regression, unreliability, was manipulated with a secondary task causing cognitive load, the overall degree of regression increased. This increase was proportionally stronger for extreme than for moderate categories.

Given the general applicability of a regression model to the frequency-judgment task, the question was what happens when some of the stimulus categories are unpacked into smaller subcategories. For example, when type-A butterflies are split into two subcategories, A_1 and A_2, marked by different colors, a small category frequency of 4 was decomposed into two extremely small subcategories with frequencies of 2 each. As the degree of regression increases with extremity, the twofold overestimation of both extremely small subcategories should result in

a larger summed estimate $f(A_1) + f(A_2)$ than the pooled estimate of $f(A_1 + A_2)$ in the nonsplit condition. Likewise, the estimated frequency of the large category D (with a frequency of 22), which is usually underestimated due to regression, can be split into two medium subcategories of size 11, which should show little regression (on a reference scale from 2 to 22). As a consequence, they should be no longer underestimated, so that the summed estimates of $f(D_1) + f(D_2)$ should again be clearly higher than the pooled estimate of $f(D_1 + D_2)$ in the nonsplit condition.

In general, the frequency illusions that resulted from the packing and unpacking of stimulus frequencies were well predicted by the general rule of regression. As expected, large quantities (e.g., frequencies) were underestimated and low quantities (frequencies) were overestimated. These discrepancies increased with greater extremity or unreliability of the available information. When the extremity of small frequencies was enhanced or the extremity of large categories was reduced (by splitting a category into smaller subcategories), this increased the subjective frequency of the entire category. The strength of category-split effects increased when cognitive load reduced the reliability of frequency assessment.

Unkelbach and Fiedler (2011) showed that regression-based effects on frequency judgments can carry over to evaluative judgments. The mere-exposure effect is the classic finding that liking for a target or stimulus increases with the frequency with which it has been encountered (Bornstein, 1989; Zajonc, 1968). Replicating this effect, with types of butterflies, Unkelbach and Fiedler found that frequently presented types were rated as more attractive than infrequently presented butterflies. Critically, however, this exposure effect varied with subjective frequencies of presentation, which, in turn, were subject to differential regression. As expected, cognitive load produced greater regression and thus smaller effects of exposure on liking. Moreover, exposure effects became smaller when stimulus categories were split into small or very small subcategories.

CONCLUSION

Despite the reputation of regression as an artifact, a noisy variance component, and a source of uncontrolled error, we have considered regression as a fascinating and enlightening property of the empirical world. Although stochastic in origin, regression effects are bound by well-understood rules that can turn unsystematic error into systematic bias. Analyzing and understanding these lawful manifestations of regression leads to alternative explanations for a host of prominent phenomena in judgment and decision making. Once we theorize in terms of regression as a model with two clearly specified determinants, extremity and unreliability, we encounter novel predictions that cannot be derived from any other theoretical approach.

We have illustrated our core message with reference to four classes of judgment biases, reflecting four different ways in which regression turns into systematic variance. The first class of phenomena includes the most prominent examples of the "regression trap" in the methodological literature. These examples typically point to illusions in the researcher's mind and they refer to misinterpretations of changes in longitudinal designs. When interventions are not initiated in a randomly determined moment but in a crisis or in an extraordinary moment, then any subsequent

improvement or return to that baseline does not justify the attribution of a separate cause (e.g., the intervention). Such changes across repeated measures may merely reflect the operation of regression to a general trend or overall mean.

The remaining three classes of phenomena are not included in classical treatises of regression artifacts. They rather reflect the systematic application of regression conceived as a theoretical model of judgment biases in several prominent paradigms. Thus, the second class originates in a critical reanalysis of the data from conditional inference studies. Conditional inferences of one variable from given levels of another variable can be misleading because regression is bidirectional. Accuracy is a regressive function of confidence but confidence is also a regressive function of accuracy. The same judgments can be overconfident and underconfident, depending on whether accuracy is conditionalized on confidence, or confidence on accuracy.

The third scenario is differential regression. Judgments and evaluations of two equivalent targets (e.g., two equally smart students; equally positive groups) will differ when the targets cannot be assessed with the same reliability, due to unequal sample size, motivation, or distance of observations. Fourth, when large target categories are decomposed into smaller subcategories or, conversely, when small categories are merged into larger superordinate categories, the summation of several regression effects can produce strong and systematic biases. Subadditivity arises from the general rule that large quantities tend to be underestimated whereas small quantities tend to be overestimated.

Regression effects can account for many well-known judgmental biases that have been traditionally explained in terms of extraneous causes. The preferred causal explanations for such biases involve intrapsychic constructs, such as stereotypes, wishful thinking, limited memory capacity, or heuristic as opposed to deliberative reasoning. Even though these factors may contribute to some of the biases some of the time, regression alone is a sufficient condition for all these phenomena to occur. For the claim that cognitive or motivational factors produce bias to be compelling, it must be demonstrated that the bias goes beyond what regression can explain. As such a research methodology has rarely been used, the canonical findings conflate to an unknown degree the impact of regression and other causal influences.

Why is there such resistance to regression as an explanatory construct? We suspect that probabilistic thinking in general is counterintuitive and that this remains true for trained scientists (Goldberg, 1991; Rulon, 1941). Formal models of probability did not appear until late in the Renaissance (Gigerenzer, Swijtink, Porter, Daston, Beatty, & Krüger, 1989). Regression is counterintuitive because it is anticyclical; it points in the direction opposite to what appear to be current trends.

In addition to these psychological barriers, theorists and researchers have incentives to explain behavior in causal terms. In experimental science, researchers take it as their mission to isolate unique causal relationships. The ability to detect causal relations is the stated distinction of the experimental method. To explain systematic biases (or other behavioral effects) with variation that is at its root random must seem like a failure to the experimentalist. Hence, we may read and understand the regression critique of overconfidence, and yet continue to believe

that overconfidence comes from the enhancement of the ego. We may comprehend the lesson that the stock market behaves regressively, while continuing to invest in stocks that were recent winners. No one has refuted the claim—truism even—that the devaluation of minorities in the illusory-correlation paradigm can reflect regression alone, but textbooks continue to broadcast the view that a memory bias toward rare (i.e., negative) behavior in rare people (i.e., minorities) is the necessary and sufficient source of this effect.

We believe that it is necessary to overcome the resistance to abstract and counterintuitive constructs in theoretical reasoning. Abstract theoretical concepts like regression can expand knowledge and stimulate bold new hypotheses. These concepts can also be translated into vivid and comprehensible theoretical metaphors. One such metaphor portrays regression as a sort of informational erosion. To the extent that the geological ground is weak and "unreliable," and to the extent that the geological surface varies in extremity of altitude, erosion will be manifested in shrinking mountain tops and rising valley floors.

Another physical metaphor, related to the concept of potential energy, construes the multivariate empirical world as a bundle of intertwined rubber strings. Depending on the direction and strength of all kinds of varying extraneous forces that impinge on the world, the individual rubber strings (i.e., variables) are stretched to varying degrees (i.e., contaminated with error variance). As a consequence, the true length of the rubber strings will be overestimated to the extent that they are exaggerated at the time of measurement. The same rubber strings can be expected to shrink on the next occasion, when external forces are eliminated or operating in different directions.

Whatever metaphor of the regression law turns out to be most intelligible and appropriate as an explanatory construct, we believe it will eventually further our theoretical and empirical knowledge. It will not only explain familiar phenomena in human judgment but also help discover new phenomena that can be derived from mental simulations or computer simulations of the regression model proper. Last but not least, systematic analyses of the conditions under which regression is minimized can assist researchers in devising efficient debiasing strategies and remedies to irrational illusions in judgment and decision making.

AUTHOR NOTE

Correspondence regarding this chapter should be addressed to Klaus Fiedler at: kf@psychologie.uni-heidelberg.de. The research underlying the present paper was supported by a Koselleck grant of the Deutsche Forschungsgemeinschaft awarded to Klaus Fiedler.

NOTES

1. Even when the standard regression model assuming a stable mean is not met in the stock market example, regression often accounts for the lion's share of outcome variance (Hubert, 1999).
2. Following a suggestion by Constantine Sedikides, these authors introduced the term *autoverification effect*.

3. The formula provided by Spearman (1910) and Brown (1910), $R = nr / (1 - r + nr)$ shows the increase in the overall reliability of a test, R, as a function of n, the number of items or responses, and the average reliability of a single item r. This function is negatively accelerated. To achieve a constant increment in reliability, the increment in sample size has to be larger for large than for small n. This is so because increasing n by a factor $k > 1$ increases the entire numerator but only one summand in the denominator by k.

REFERENCES

Alicke, M. D., & Sedikides, C. (2009). Self-enhancement and self-protection: What they are and what they do. *European Review of Social Psychology, 20*, 1–48.

Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research, 1967–1987. *Psychological Bulletin, 106*, 265–289.

Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology, 3*, 276–322.

Campbell, D. T. (1996). Regression artifacts in time-series and longitudinal data. *Evaluation and Program Planning, 19*, 377–389.

Campbell, D. T., & Kenny, D. A. (1999). *A primer on regression artifacts*. London: Guilford Press.

Dawes, R. M., & Mulford, M. (1996). The false consensus effect and overconfidence: Flaws in judgment or flaws in how we study judgment? *Organizational Behavior and Human Decision Processes, 65*, 201–211.

Dovidio, J. F., & Gaertner, S. L. (2010). Intergroup bias. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (5th ed., Vol. 2, pp. 1084–1121). Hoboken, NJ: Wiley.

Epley, N., & Whitchurch, E. (2008). Mirror, mirror on the wall: Enhancement in self-recognition. *Personality and Social Psychology Bulletin, 34*, 1159–1170.

Erev, I., Wallsten, T. S., & Budescu, D. V. (1994). Simultaneous over- and underconfidence: The role of error in judgment processes. *Psychological Review, 101*, 519–527.

Eysenck, H. J. (1965). *Fact and fiction in psychology*. Baltimore, MD: Pelican.

Fiedler, K. (1991). The tricky nature of skewed frequency tables: An information loss account of distinctiveness-based illusory correlations. *Journal of Personality and Social Psychology, 60*, 24–36.

Fiedler, K. (1996). Explaining and simulating judgment biases as an aggregation phenomenon in probabilistic, multiple-cue environments. *Psychological Review, 103*, 193–214.

Fiedler, K. (2000). Illusory correlations: A simple associative algorithm provides a convergent account of seemingly divergent paradigms. *Review of General Psychology, 4*, 25–58.

Fiedler, K. (2002) Frequency judgment and retrieval structures: Splitting, zooming, and merging the units of the empirical world. In P. Sedlmeier &. T. Betsch (Eds.), *Frequency processing and cognition* (pp. 67–87). Oxford, UK: Oxford University Press.

Fiedler, K. (2011). Voodoo correlations are everywhere—Not only in neuroscience. *Perspectives on Psychological Science, 6*(2), 163–171.

Fiedler, K., & Armbruster, T. (1994). Two halfs may be more than one whole: Category-split effects on frequency illusions. *Journal of Personality and Social Psychology, 66*, 633–645.

Fiedler, K., Freytag, P., & Unkelbach, C. (2007). Pseudocontingencies in a simulated classroom. *Journal of Personality and Social Psychology, 92*, 665–667.

Fiedler, K., Unkelbach, C., & Freytag, P. (2009). On splitting and merging categories: A regression account of subadditivity. *Memory & Cognition, 37*, 383–393.

Fiedler, K., Walther, E., Freytag, P., & Plessner, H. (2002). Judgment biases in a simulated classroom—A cognitive–environmental approach. *Organizational Behavior and Human Decision Processes, 88*, 527–561.

Fiedler, K., Walther, E., & Nickel, S. (1999). The autoverification of social hypothesis: Stereotyping and the power of sample size. *Journal of Personality and Social Psychology, 77*, 5–18.

Furby, L. (1973). Interpreting regression toward the mean in developmental research. *Developmental Psychology, 8*, 172–179.

Galton, F. (1886). Regression toward mediocrity in hereditary stature. *Journal of the Anthropological Institute of Great Britain and Ireland, 15*, 246–263.

Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., & Krüger, L. (1989). *The empire of chance: How probability changed science and everyday life*. New York: Cambridge University Press.

Goldberg, L. R. (1991). Human mind versus regression equation: Five contrasts. In D. Cicchetti & W. M. Grove (Eds.), *Thinking clearly about psychology: Essays in honor of Paul E. Meehl* (Vol. 1, pp. 173–184). Minneapolis: University of Minnesota Press.

Hamilton, D. L., & Gifford, R. K. (1976). Illusory correlation in interpersonal perception: A cognitive basis of stereotypic judgments. *Journal of Experimental Social Psychology, 12*, 392–407.

Hubert, M. (1999). The misuse of past-performance data. In L. E. Likson & R. A. Geist (Eds.), *The psychology of investing* (pp. 147–157). New York, NY: Wiley.

Jennions, M. D., & Møller, A. P. (2002). Relationships fade with time: A meta-analysis of temporal trends in publication in ecology and evolution. *Proceedings of the Royal Society London B, 269*, 43–48.

Juslin, P., Winman, A., & Olsson, H. (2000). Naive empiricism and dogmatism in confidence research: A critical examination of the hard–easy effect. *Psychological Review, 107*, 384–396.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(2), 263–291.

Klar, Y., & Giladi, E. (1997). No one in my group can be below the group's average: A robust positivity bias in favor of anonymous peers. *Journal of Personality and Social Psychology, 73*, 885–901.

Klauer, K., & Meiser, T. (2000). A source-monitoring analysis of illusory correlations. *Personality and Social Psychology Bulletin, 26*, 1074–1093.

Klayman, J., & Ha, Y. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94*, 211–228.

Krueger, J. (1998). Enhancement bias in the description of self and others. *Personality and Social Psychology Bulletin, 24*, 505–516.

Krueger, J., & Mueller, R. (2002). Unskilled, unaware, or both? The better-than-average heuristic and statistical regression predict errors in estimates of own performance. *Journal of Personality and Social Psychology, 82*, 180–188.

Krueger, J. I., & Wright, J. C. (2011). Measurement of self-enhancement (and self-protection). In M. D. Alicke & C. Sedikides (Eds.), *Handbook of self-enhancement and self-protection* (pp. 472–494). New York: Guilford.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*, 1121–1134.

Langer, E. (1975). The illusion of control. *Journal of Personality and Social Psychology, 32*, 311–328.

McGarty, C., Haslam, S. A., Turner, J. C., & Oakes, P. J. (1993). Illusory correlation as accentuation of actual intercategory difference: Evidence for the effect with minimal stimulus information. *European Journal of Social Psychology, 23*, 391–410.

McNemar, Q. (1940). A critical examination of the University of Iowa studies of mental influences upon the IQ. *Psychological Bulletin, 37*, 63–92.

Moore, D. A., & Small, D. A. (2007). Error and bias in comparative judgment: On being both better and worse than we think we are. *Journal of Personality of Social Psychology, 92*, 972–989.

Morewedge, C. K., Gilbert, D. T., Keyzar, B., Berkovits, M. J., & Wilson, T. D. (2007). Mispredicting the hedonic benefits of segregated gains. *Journal of Experimental Psychology: General, 136*, 700–709.

Mullen, B., & Johnson, C. (1990). Distinctiveness-based illusory correlations and stereotyping: A meta-analytic integration. *British Journal of Social Psychology, 29*, 11–28.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*, 175–220.

Ross, L., & Nisbett, R. E. (1991). *The person and the situation*. New York: McGraw-Hill.

Rottenstreich, Y., & Tversky, A. (1997). Unpacking, repacking, and anchoring: Advances in support theory. *Psychological Review, 104*, 406–415.

Rulon, P. J. (1941). Problems of regression. *Harvard Educational Review, 11*, 213–223.

Sedlmeier, P. (1999). *Improving statistical reasoning: Theoretical models and practical implications*. Mahwah, NJ: Lawrence Erlbaum Associates.

Snyder, M. (1984). When belief creates reality. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 18, pp. 247–305). New York: Academic Press.

Soll, J. B. (1996). Determinants of overconfidence and miscalibration: The roles of random error and ecological structure. *Organizational Behavior and Human Decision Processes, 65*(2), 117–137. doi:10.1006/obhd.1996.0011.

Spearman, C. C. (1910). Correlation calculated from faulty data. *British Journal of Psychology, 3*, 271–295.

Stelzl, I. (1982). *Fehler und Fallen der Statistik [Errors and traps of statistics]*. Bern, Switzerland: Huber.

Stigler, S. (1999). *Statistics on the table: The history of statistical concepts and methods*. Cambridge, MA: Harvard University Press.

Thompson, S., Armstrong, W., & Thomas, C. (1998). Illusions of control, underestimations, and accuracy: A control heuristic explanation. *Psychological Bulletin, 123*, 143–161.

Thorndike, E. L. (1924). The influence of chance imperfections on measures upon the relationship of initial score to gain or loss. *Journal of Experimental Psychology, 7*, 225–232.

Thorndike, R. L. (1942). Regression fallacies in the matched groups experiment. *Psychometrika, 7*, 85–102.

Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review, 101*, 547–567.

Unkelbach, C., & Fiedler, K. (2011). *Regression-based frequency illusions and the mere exposure effect*. Manuscript submitted for publication.

Weisz, J. R., McCarty, C. A., & Valeri, S. M. (2006). Effects of psychotherapy on depression in children and adolescents: A meta-analysis. *Psychological Bulletin, 132*, 132–149.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure [Monograph supplement]. *Journal of Personality and Social Psychology, 9*(2, Pt. 2), 1–27.