Running head: PERFORMANCE EVALUATION

Correspondence Bias in Performance Evaluation:
Why Grade Inflation Works

Don A. Moore

Samuel A. Swift

Zachariah S. Sharek

Francesca Gino

Carnegie Mellon University

Corresponding author:
Don Moore
CMU/Tepper
5000 Forbes Avenue
Pittsburgh, PA 15213
don.moore@alumni.carleton.edu
Phone: 412-268-5968
FAX: 412-268-7345

The authors appreciate the support of National Science Foundation Grant SES-0451736 and a SURG research grant from Carnegie Mellon University. The authors owe Hart Blanton a profound debt of gratitude for setting them straight on analyzing the data from the first study. Thanks also to Mark Fichman, Bill Klein, and Justin Kruger for helpful comments. Thanks to Lauren DeVito, Bill Mangan, and Jessica Wisdom for help with data collection. Address correspondence to don.moore@alumni.carleton.edu.

Correspondence Bias in Performance Evaluation: Why Grade Inflation Works

Abstract

When explaining others' behaviors, achievements, and failures, it is common for people to attribute too much influence to the individual's disposition and too little influence to the structural and situational influences impinging on the actor. Although performance is a joint function of ability and situational facilitation or impediments, dispositional inference ascribes too much to individual ability. We hypothesize that this tendency leads university admissions decisions to favor students coming from institutions with lenient grading because they will have their high performance mistaken for evidence of high ability. In four studies using both laboratory experiments and actual admissions decisions, we show that those who display high performance simply due to lenient grading or to an easy task are favored in selection. These results have implications for research on attribution, because they provide a more stringent test of the correspondence bias, and allow for a more precise measure of its size. Implications for admissions and personnel selection decisions are also discussed.

Correspondence Bias in Performance Evaluation:

Why Grade Inflation Works

Who is likely to be the more ambitious and hard-working employee—the one who was born into privilege and attended excellent and expensive private schools, or the one who, starting from the most humble beginnings, clawed her way out of the ghetto and worked to put herself through college? Hiring and promotion decisions are some of the most important ones in any organization, and they are filled with exactly this sort of difficult attribution problem. Job candidates vary with respect to both their abilities and obstacles they have had to overcome to be successful. The same is true, of course, in academia—in some fields, for instance, researchers can run a large number of small experiments and write many papers. In other fields, such as those that rely more heavily on field data, the higher cost of obtaining data usually results in fewer publications. The question we pose in this paper is whether those making the selection decisions can adequately adjust for the difficulty of success when making inferences about what performance signals about abilities.

Inferences of ability and disposition are crucial to personnel selection decisions (Staw, Bell, & Clausen, 1986; Staw, Sutton, & Pelled, 1994). Universities, when they make their admissions decisions, face the selection problem on a more structured calendar than do most other organizations. University admissions processes represent some of the most systematic and rigorous selection processes of any organization. Businesses are still willing to pay a premium to hire MBAs, whether or not what they learned in business school actually improves their productivity at work (Pfeffer & Fong, 2002), in part because university admissions serve as such an effective selection mechanism (Pfeffer, 1977; Pfeffer & Fong, 2004).

In this paper, we focus on university admissions decisions both because they are so important and because the data on admissions are so useful for testing our hypotheses. We begin with the fact that undergraduate institutions vary in their grading standards (Rojstaczer, 2003). This fact raises the question of whether those who use information about grades to assess students (such as future employers or graduate schools) use it appropriately. Do people appropriately adjust their interpretation of grades based on the leniency of grading? Making effective use of GPA information requires knowledge of the norms and standards at the institution from which the candidate comes. A 3.5 GPA at Purdue University (where the average GPA is 2.8) would distinguish a candidate from his or her peers. At the University of Michigan (where the average GPA is 3.5), however, it would be less remarkable. Applicants who come from institutions characterized by high average grades ought to have their GPAs discounted or adjusted downward. Research findings in attribution give us reason to doubt that people will be able to use this sort of discounting information effectively.

Biases in the Attribution Process

The problem of assessing the informative value of academic grades is a special case of the more general problem: How to infer the qualities of the individual (such as intellectual abilities) from behavior or outcomes (a GPA), while subtracting out the influence of the situation (leniency of grading). The solution to this problem is provided by Kurt Lewin's attributional equation: *Disposition = Behavior – Situation*. In this case, *Academic Performance = Student's GPA – Average GPA*.¹ Research suggests, however, that people do not apply this simple formula perfectly. One of the most common biases in the attribution process is the tendency to ascribe too little causal influence to the situation and therefore to attribute too much influence to the individual's disposition as a cause of behavior (Davis-Blake & Pfeffer, 1989; Gilbert, 1994;

Jones & Harris, 1967; Nisbett & Borgida, 1975; Ross, 1977). For instance, research has shown that attributions of performance by supervisors tend to gravitate toward dispositional attributions (Martin & Klimoski, 1990; Mitchell & Wood, 1980). In a different study, Ross, Amabile, and Steinmetz (1977) paired participants and randomly assigned one of them to the role of quiz master, who made up the questions, and one to the role of quiz taker, who answered them. Naturally, the quiz master knew some things that the quiz taker did not. But rather than attributing this to the quiz master's role-conferred advantage, observers reported that the quiz master was the more knowledgeable of the two. Attributions of knowledgeability were biased by an excessive belief in the correspondence between behaviors and dispositions. This is what Gilbert and Malone (1995) called the correspondence bias.

Is the Correspondence Bias Really a Bias?

In Ross, Amabile, and Steinmetz's (1977) experiment, as in other studies demonstrating the correspondence bias, it is difficult for individual participants to determine the strength of the situation precisely. On average, the quiz master is unlikely to be more knowledgeable than the quiz taker, but that is little help for the individual who must decide whether this specific quiz master is more or less knowledgeable than this particular quiz taker. Similar ambiguity about the situation confronted participants in Jones and Harris's classic (1967) study, who were asked to infer the attitudes of people forced to write an essay either for or against Fidel Castro. Participants might have wondered what compelled the essay-writer to write in praise of Castro: What exactly was the punishment for writing an essay that did not sufficiently comply with the assignment? Perhaps it should not be surprising that in these and other studies on the correspondence bias, people neglect the power of the *Situation* given the difficulties determining its strength. In order to accurately judge the strength of the situation, participants in Ross,

Amabile, and Steinmetz's experiment needed to know what proportion of questions, on average, quiz takers failed to answer correctly. If participants had this information, they would have been better able to specify the strength of the situational differences between the quiz master and quiz taker. But they did not get it. If it was impossible to determine the strength of the situation, then it becomes impossible to use it when making attributions. This fact suggests that the correspondence bias, as it has been demonstrated previously, might not actually be a bias, but rather an incomplete information problem. We address this possibility by examining if the correspondence bias persists when people have all the information they need in order to adjust their attributions of individual abilities based on the influence of the situation: participants in our experiments are given quantified information about both the behavior (i.e., grades) and the situation (i.e., grading leniency). Previous research on the correspondence bias has not tested the bias in situations where participants have knowledge of both the situation and the outcome. Furthermore, prior studies failed to specify the strength of the situation and thus left important questions unanswered about the causes for the effects they obtained.

A second contribution of our research is that we address another alternative explanation for findings of the correspondence bias, one that is related to the first. Schwarz (1996; Bless, Strack, & Schwarz, 1993) has pointed out that in experiments on the correspondence bias, researchers provide participants with better information about others' behavior than about the situation. If participants sensibly assume that the experimenter must have given them the information because it was relevant to the dispositional judgment they were being asked to make, that could account for their tendency to rely on it so heavily.

For instance, Gilbert and Jones (1986) had their participants listen to verbal expressions of attitudes that they believed another person had been compelled to read aloud. The results

showed that people made correspondent inferences, reporting that the other person must be liberal if they made a majority of liberal statements. The problem here is that it is not clear how exactly participants should have made use of the heartfelt delivery of ideology they had heard. Gilbert and Jones suggest that the answer is that participants should have ignored it entirely. However, after presenting people with compelling and vivid (if misleading) information, it seems unfair to accuse them of behaving irrationally when they pay attention to it.

In the past, the correspondence bias has been explained as a result of people's underestimation of the power of the *Situation* (as in Ross et al., 1977) and overestimation of the informative value of *Behavior* (as in Gilbert and Jones, 1986). Yet prior research has not examined contexts in which people know both *Situation* and *Behavior* with unambiguous certainty. Although prior theorizing (e.g., Gilbert, 1994) would imply that the correspondence bias should go away under these circumstances, prior research cannot tell us whether it will—whether underestimation of the *Situation* or overestimation of *Behavior* are necessary conditions for producing the correspondence bias. Perhaps when it is so clear and so simple and when people have plenty of time and cognitive resources to consider it, then the attributional problem becomes one of trivial simplicity. In this paper, we advance theory and research by testing this possibility. The problem of computing *Disposition* when people know *Behavior* and *Situation* precisely ought then to come down to simple subtraction. Can the correspondence bias persist, even in this context?

Hypothesis

The primary hypothesis we investigate is that absolute performance will be insufficiently discounted based on the ease of the task. Specifically, the correspondence bias would predict that raw GPAs will be taken as evidence of academic performance and not sufficiently adjusted

to account for the ease with which those grades were earned. Put another way, an applicant's absolute GPA will have a stronger influence on admission decisions than will the ease with which those grades were achieved. Rationally, evaluations of an applicant's prior academic performance should rely primarily on two things: (1) the rigor or quality of the institution or program of study and (2) the individual's relative performance compared to others in that same program. Our studies control for the first consideration and vary two aspects of second: absolute performance (GPA) and the ease of obtaining a high score (average grades at undergraduate institution). The easier the task, the less impressive high performance becomes. If evaluations were being made optimally, then leniency of grading (as measured by the average grade at undergraduate institution) should be given a decision weighting equal in size and opposite in sign from that given to each candidate's GPA.

To draw an analogy, assume your goal is to pick the tallest players for your basketball team. In a desperate attempt to improve their chances of making the team, some of those trying out have worn elevator shoes. If you know only the player's total height (with shoes) and the height of the shoes, then those two measurements should be weighed equally and oppositely in determining the player's shoeless height: Each inch contributed by the shoe will reduce by one inch the player's size when the shoes are removed. Instead, we hypothesize that the positive effect of individual performance (e.g., height) on evaluations will not be matched by the discounting effect of situational factors (e.g., elevator shoes). In the context of grades and admission, our hypothesis predicts that people will favor those from institutions with lenient grading because absolute GPA will be weighed heavily in evaluations of applicants, but average grades at the institution will not be used sufficiently to discount it. In other words, those who show up in elevator shoes will be more likely to make the team.

We opted to first establish the existence in field data of the effect we hypothesize. Study 1 examines the relative effects of undergraduate GPA and the undergraduate institution's average grades on graduate school admissions decisions. Studies 2 and 3 take the graduate school admissions decision into the laboratory in order to better establish causal order and experimental control. Study 4 is another lab study that replicates the same decision problem outside the context of GPAs and admissions decisions. All four studies are consistent in showing that nominal performance is too readily taken at face value without being discounted by obvious situational influences, even when the effect of those situational influences is obvious and is easily quantified.

STUDY 1

Study 1 tests the degree to which the leniency of undergraduate grades affects graduate school admissions decisions. We test whether, controlling for the other features of applicants, the average grades at an applicant's undergraduate institution carries any weight in admissions decisions.

Method

We obtained admissions data from two selective masters programs at large private universities. These data include each applicant's undergraduate GPA and alma mater, as well as a number of useful control variables such as age, gender, race, national citizenship, years of work experience, standardized test scores (in this case, the GMAT test), and interview performance. We collected data on the quality of each applicant's undergraduate college or university from rankings conducted by U.S. News & World Report (2005).

While the original applicant pool totaled 3146, many of these applicants had to be excluded from the analysis. Applicants from foreign institutions were excluded because foreign

grading systems are different from those used by American institutions. Applicants from U.S. institutions who did not report their GPAs were also excluded. The remaining 1683 applicants applied from 407 different undergraduate institutions. We attempted to collect data on average grades for all 407 institutions and were able to obtain data on the average GPA from 182 of these institutions covering 59% of applicants. The data on average grades we were able to collect came from a variety of sources. Average grade data from 35 schools were obtained from the world wide web: www.gradefinflation.com (Rojstaczer, 2003). However, this web site is far from comprehensive. We contacted the remaining institutions directly by email, or when that failed to produce a response, by telephone. Many declined to provide us with information about average grades.² The institutions from which we were able to obtain average grade information did not differ systematically from the rest with respect to private/public status, size, tuition, U.S. News ranking, or entrance SAT score.

Results and Discussion

The means, standard deviations, and bivariate correlations between variables are shown in Table 1. There is, not surprisingly, a positive relationship between applicants' undergraduate grades and admission decisions: Those who were admitted had higher GPAs ($M = 3.40$) than those who were not ($M = 3.24$), $t(1681) = 63.79$, $p < .001$. Those admitted also tended to come from institutions with higher average grades ($M = 3.23$, $SD = .18$) than those who were not admitted ($M = 3.19$, $SD = .18$), $t(988) = 3.46$, $p = .001$. It is possible that the positive effect of institutional GPAs is because better universities have higher average grades. Indeed, average grades tend to be highest at the most selective universities ($r = .51$), suggesting that average absolute GPA may in fact be correlated with student ability. A more complete picture, then, is provided by a full multiple regression.

In order to examine which factors predict admission, we conducted a binary logistic regression on all 1683 cases. The results of this model are shown in Table 2 as Model 1. Note that the logistic regressions in Table 2 omit some institution-level control variables (i.e., public/private status, tuition, and entrance SAT score) that are strongly correlated with our measure of institution quality. Their inclusion made the coefficient associated with institution quality unstable, and none of them emerged as significant themselves. Their omission has no material effect on the key variables of interest, namely the applicant's GPA and the average GPA at the applicant's alma mater.

Of course, a high undergraduate GPA increases an applicant's chances of being admitted ($B = 1.38, SE = .18, p < .001$). Model 2 adds average grades at each institution as an independent variable, and thus includes only the 990 cases for which we were able to obtain data on average grades. The results are consistent with our hypothesis. The results show that average grades at undergraduate institution were not a significant influence on admission decisions, ($B = -.55, SE = .6, p = .36$). If higher average grades had been used to discount the GPAs of applicants from schools with high average grades, then it should have had a significant negative effect on admission. It did not.³ The negative coefficient on grading leniency was only 36% of the size of the positive coefficient on GPA.

Before concluding that grading leniency had no effect on the interpretation of GPAs, we also wanted to test for possible higher-order effects. For instance, it is possible that while lenient grading did not have a direct negative effect on admissions decisions, it did nevertheless affect the interpretation of GPAs. In order to test this hypothesis, we conducted a third regression that added an interaction term for *GPA * average GPA at undergraduate institution*. This interaction term was not significant ($B = -.11, SE = .21, p = .61$). Similarly, neither the *quality of*

*undergraduate institution * average GPA at undergraduate institution* interaction ($B = -.92$, $SE = .96$, $p = .34$) nor *GPA * quality of undergraduate institution* ($B = .73$, $SE = .49$, $p = .13$) were significant predictors of graduate school admission. It would appear that grading leniency simply did not have much affect on admissions decisions.

We argue that the tendency for admissions decisions to ignore grading norms at the undergraduate institution is explainable by attributional processes that take performance at face value and fail to discount it using situational influences. The result is that applicants from schools with high average grades are evaluated more positively simply because their GPAs are higher. However, we ought to be cautious interpreting what is essentially a null finding—namely the lack of a significant negative effect of lenient grading on admissions decisions. Moreover, the field data do not allow us to rule out two viable alternative explanations for our results. First, it might not be fair to assume that admissions staffs have perfect information about average grades at all other academic institutions. After all, these data were costly for us to collect. Without the data, it would be difficult for admissions departments to use the information to discount grades appropriately, and they would simply have to rely on absolute GPAs as useful (if imperfect) measures of academic performance. On the other hand, this information is critical to their ability to interpret applicants' GPAs, and admissions offices are likely to have better ways of obtaining these data than did we. A failure to demand it, in and of itself, suggests a failure to appreciate its value.

Second, it is possible that lenient grading captures something important about the undergraduate institution. While we considered the roles of institution quality, public/private status, tuition, and student quality as control variables, it is possible that lenient grading covaries with other important features of institutions that are not captured in our control variables. Since

the field data do not allow us to rule out this concern, we turn to the experimental laboratory to provide us with a more controlled test of our hypothesis. Our laboratory studies help clarify how information is used in making dispositional inferences, and help answer the question of whether the results of Study 1 are due to a lack of information about average grades at different institutions.

STUDY 2

This experiment put participants in the role of an admissions decision maker and presented them with information about specific candidates' performance (GPA) as well as an indication of the distribution from which the GPA came (college average GPA). We manipulated these two factors in a 3×3 design. Applicants came from colleges that were of similar quality and selectivity, but with average grades that were either high (average GPA of 3.6), medium (3.0), or low (2.4). This manipulation was crossed with a manipulation of the candidates' GPAs relative to their classmates: Candidates had GPAs that were high (.3 above their school's average), medium (at their school's average), or low (.3 below their school's average).⁴ We expected that, as in Study 1, applicants' nominal GPAs would exert a strong influence on admissions decisions, but that the effect of the grading norms at their alma maters would not be as strong. In terms of our experimental manipulations, that led us to expect that being above average would have a positive effect on the probability of being admitted. Our more interesting hypothesis is that the school's average GPA would have a significant positive effect on the probability of admission: Candidates from colleges with high average grades would be more likely to be admitted. In other words, we expect that people will not sufficiently discount high grades that are due to lenient institutional grading practices.

Method

Participants. Participants were 55 undergraduates at a university in the United States participating in exchange for course credit in their introductory business courses.

Procedure. Participants were given the following instructions:

“In this exercise, you will be playing the role of a member of the admissions committee at a selective MBA program. You are selecting students who would like to obtain masters degrees in business administration. Your most important goal is to select the best candidates from among the applicants. In general, you usually have space to admit about half the applicants.

“You will see the applications of nine hypothetical students. The set of applicants that you will review all graduated from colleges of similar quality and selectivity. Please review each applicant carefully in order to assess the quality of their prior academic performance in college. Please review one candidate at a time. Answer the questions about each candidate before turning the page to read about the next candidate.”

Information about each candidate included: GPA, the average GPA at the institution from which they obtained their undergraduate degrees, and grades in the last 10 classes they took. These classes were listed for each candidate. The candidate's grade in each class was shown, as was the average grade in the class. The candidate's grades in the 10 classes had the same mean as the candidate's overall GPA, and the average grades in each of the courses, when averaged, had the same mean as the undergraduate institution overall. In order to highlight each candidate's relative standing, the difference between his or her GPA and the average for the college was also specifically shown. This list of classes was counterbalanced between subjects

so that each list of classes had an equal chance of appearing in any of the nine experimental conditions.

For each candidate, participants were asked to (1) evaluate how successful the candidate had been in college on a 7-point scale, anchored at 1 (“*very unsuccessful*”) and 7 (“*very successful*”), and (2) report how likely they would be to admit them (as a numerical probability between 0% and 100%). After evaluating all nine candidates, they were asked to look back through the set and admit only four of the nine. In sum, for each candidate, each participant provided three ratings: (1) a rating of prior success, (2) an estimated probability of admission, and (3) an actual admission decision.

Participants were randomly assigned to one of nine randomly-determined order conditions in a Latin-squares design such that each candidate’s position in the order was balanced. Names of the nine fictional colleges and course lists were counterbalanced across manipulations.

Results and Discussion

The three ratings of each candidate correlated strongly with each other (all r ’s above .6). Therefore, the three ratings of each candidate were standardized (by converting them to z-scores) and averaged to form a single measure (alpha reliability = .86).

This overall assessment was then subjected to a 3 (GPA relative to average) $\times$ 3 (average grade at undergraduate institution) within-subjects ANOVA. Naturally, the results show a main effect of relative GPA, $F(2, 108) = 333.84, p < .001, \eta^2 = .86$. Those with above-average GPAs were dramatically more likely to be offered admission ($M = 82\%$) than were those with below-average GPAs ($M = 13\%$). As hypothesized, the results also show a significant main effect for average grade at undergraduate institution, $F(2, 108) = 96.35, p < .001, \eta^2 = .64$. Consistent

with our expectations, candidates from institutions with lenient grading were more likely ($M = 62\%$) to be offered admission than were candidates from schools with strict grading ($M = 22\%$). The results also reveal a GPA $\times$ leniency interaction effect, $F(4, 216) = 6.44, p < .001, \eta^2 = .11$. This interaction appears to be a result of the fact that the effect of lenient grading was strongest for candidates from institutions with medium average grades, although this interaction effect seems to hold only for admissions decisions and not for the other two dependent measures (see Table 3).

In order to compare the results of Study 2 with those of Study 1, we conducted a binary logistic regression in which the probability of being offered admission was the dependent variable and each candidate's GPA and average grades at undergraduate institution served as independent variables. The results reveal that both are significant predictors. Unsurprisingly, absolute GPA had a significant positive relationship with the rated probability of admission ($B = 7.22, SE = .66, p < .001$). More interestingly, coming from an institution with lenient grading hurt applicants ($B = -4.85, SE = .57, p < .001$), but the size of this effect was significantly smaller in magnitude than the effect of raw GPA, $\chi^2(1) = 65.30, p < .0001$. In other words, participants did discount a high GPA somewhat if it came from an institution with lenient grading, but this effect was too small to undo the strong positive effect of having a higher GPA. The size of the negative effect of grading leniency was only 67% the size of the positive effect of having a higher GPA.

The results of Study 2 are consistent with the hypothesis that absolute GPAs are taken as direct evidence of prior academic performance and are not appropriately discounted by the ease with which those grades were earned. These results replicate and extend prior work on comparative judgment, which has shown that feedback regarding absolute performance has a

powerful influence on comparative judgment. For instance, Moore and Kim (2003) gave participants either simple or difficult trivia quizzes. When participants had taken a simple trivia quiz, they knew that they had gotten most of the questions correct; they also believed themselves to be above average and were willing to bet on beating a randomly selected opponent. When the test was difficult, participants believed they were below average and were not interested in betting. While the results of Study 2 are consistent with the findings of Moore and Kim (2003), the present results are more striking because participants in Study 2 had exactly the information they needed to solve the attributional equation. Moore and Kim's participants only had imperfect information about their own performances and had to guess others' performances. Despite having perfect information about both absolute and relative performance of all the applicants in the set, participants in Study 2 were clearly overly influenced by the absolute performance information at their disposal.

The clarity of information we gave our participants is also notable because it diverges from the usual methods used to study the correspondence bias in prior research. The fact that we provided our participants with clear quantification of both people's behavior (their GPAs) and the situation that gave rise to that behavior (the average grades at that institution), means that we can make stronger claims about our results than can prior research. First, the correct answer is clearer, in the sense that GPA and grading leniency should have been equally and oppositely weighted. The fact that they are not allows us to pinpoint exactly how it is that our participants' decisions deviate from the optimal, and how much this matters. In Study 2, grading leniency received a decision weighting that was 67% the size of the weighting of GPA. In Study 1, it was only 36% the size. We obtain two more estimates of this discrepancy from Studies 3 and 4.

STUDY 3

One potential concern regarding the internal validity of Study 2 is that there was simply greater variance in the GPAs of the nine candidates than there was in the average grades at their home institutions: GPA's ranged from 2.1 to 3.9 while average grades ranged from 2.4 to 3.6. On the one hand, this reflects the real state of the world because there will always be more variation in the grades of individual students than in the average grades of groups of students. But on the other hand, this experimental design may have "loaded the dice" in favor of finding an effect of absolute GPA because more extreme independent variables are likely to have stronger effects on the dependent variables. Study 3 sought to address this concern by equalizing the variance in the two independent variables.

In the first two studies, we observed a positive effect of high average grades: Average students from institutions with lenient grading were more likely to be admitted than were average students from institutions with strict grading. We expect that effect to be replicated in this third study. Note, however, that because we hold absolute GPAs constant, as long as participants engage at all in the sensible tendency to discount grades that come from institutions with lenient grading, the paradigm in this third study will produce a negative overall effect for grade leniency. The more interesting question will be the comparison between its size and the size of the effect for absolute GPA.

Method

Participants were 45 undergraduates at a university in the United States participating in exchange for course credit.

The task for Study 3 was the same as the previous experiment: to evaluate nine candidates and eventually decide which four to admit. The one difference was in the GPA

numbers. Participants in this experiment saw applicants who had GPAs that were high (3.6), medium (3.0) or low (2.4). They came from undergraduate institutions at which average grades were either high (3.6), medium (3.0), or low (2.4).

Participants again rated each of the nine candidates' success in college and likelihood that they would be accepted. As in Study 2, after reviewing the field individually, they then decided which four to accept into their MBA program.

Results and Discussion

As in Study 2, the three ratings of each candidate were standardized and averaged to form a single measure (alpha reliability = .81).

This overall assessment was then subject to a 3 (GPA) $\times$ 3 (average grade at undergraduate institution) within-subjects ANOVA. The results demonstrate the expected main effect of GPA, $F(2, 88) = 492.8, p < .001, \eta^2 = .92$. Those with high GPAs were dramatically more likely to be offered admission ($M = 90\%$) than were those with low GPAs ($M = 1\%$). The results also show a more modest but significant main effect for grade leniency, $F(2, 88) = 29.73, p < .001, \eta^2 = .40$. Candidates from institutions with lenient grading were less likely ($M = 30\%$) to be offered admission than were candidates from schools with strict grading ($M = 54\%$). The results also reveal a GPA $\times$ leniency interaction effect, $F(4, 176) = 2.72, p = .03, \eta^2 = .06$. This interaction appears to be a result of the fact that the effect of lenient grading was strongest for candidates from institutions with medium average grades, although this interaction effect seems to hold only for admissions decisions and not for the other two dependent measures (see Table 4).

Again, in order to compare the results of Study 3 with those of Studies 1 and 2, we conducted a binary logistic regression in which the probability of being offered admission was

the dependent variable and each candidate's GPA and average grades at undergraduate institution served as independent variables. The results again reveal that both are significant predictors. Unsurprisingly, raw GPA had a significant positive relationship with the probability of admission ($B = 5.48$, $SE = .52$, $p < .001$). More interestingly, coming from an institution with lenient grading hurt applicants ($B = -1.90$, $SE = .35$, $p < .001$), but the magnitude of this effect was significantly smaller than the effect of raw GPA, $\chi^2(1) = 53.68$, $p < .0001$. The discounting effect was only 34% the size of the effect of a higher GPA.

Again, people discounted high GPAs somewhat when they came from institutions with lenient grading, but this effect was too small to undo the strong positive effect of having a higher GPA. To be specific, an increase of one point in an applicant's GPA increased his or her rated probability of admission by 34.9%. But if that one point increase came solely because everyone at that institution got higher grades, it still increased the rated probability of admission by 27.3%. An increase in grading leniency of an entire grade point at the undergraduate institution only decreased the rated probability of admission by 7.59%.

The results of Study 3 replicate the key finding of the first two studies. Holding relative performance constant, admission candidates were evaluated more positively when they came from institutions with higher grades. Candidates with average grades from a college with an average GPA of 3.6 were admitted by 76% of participants. Candidates with average grades from a college with an average GPA of 2.4 were selected by 4% of participants.

STUDY 4

One potential concern regarding Studies 2 and 3 is that if people believe that high average grades are correlated with desirable features of a college or its graduates, despite our assurance that the institutions did not differ with regard to quality and selectivity, the tendency to

favor graduates of institutions with high average grades makes sense. In order to rule out this explanation for our findings, we elected to conduct a fourth study outside the domain of university admissions decisions. Instead, participants in Study 4 were asked to imagine that they had to select members for a “quiz bowl” trivia team. They reviewed the prior test performances of ten applicants, five of whom had taken an easy test and five of whom had taken a difficult test. Both tests were on the subject of U.S. states and geography. Our hypothesis was that those who had high scores because they had taken the easy test would, like those who have come from institutions with lenient grading, be more likely to be selected.

Method

Participants were 71 undergraduates at a university in the United States participating for money.

The task for Study 4 was similar to Studies 2 and 3 to evaluate ten candidates and eventually decide which five to select. However, participants in this experiment were asked to select the candidates they thought would perform above average on a third quiz which was shown to participants at the time of selection:

“In this study, we are interested in your ability to predict the performance of others. You will now see the scores of 10 people who took one of two quizzes. For five of the people, you will see their scores on the first quiz. For the other five, you will see their scores on the second quiz. For each of the 10 contestants, their correct answers are marked with a check and their incorrect answers are marked with an X. For each of the ten contestants, we will ask you to estimate their knowledgeability about US states and geography. All ten of these people also took a third quiz on the same topic of US geography. After examining 10 quizzes, we will ask you to identify the 5 people you think are most likely to perform well on the third quiz. This third quiz

was the same for all 10 contestants. You will earn \$2 today for each contestant you pick whose score is in the top half of the performers on the third quiz. Therefore, if you correctly pick the 5 top scorers, we will reward you with \$10 in cash for your performance. If the five contestants you pick are the five worst performers on the third quiz, you will not receive earn any additional money for this exercise”.

Unlike Studies 2 and 3, this experiment also manipulated experience with the task. People in the experience condition were given an additional page with instructions at the beginning of the study: *“Your first task in this study is to take two different trivia quizzes. Your goal is to answer as many questions correctly as you can, using your memory alone. You may not consult other people or information sources other than your own memory. Good luck!”*

For each of the ten applicants, participants saw a completed quiz which included candidates’ answers marked as correct or incorrect. These quizzes were actual completed quizzes from participants in a previous pilot study. Participants saw quizzes from five candidates who had taken a difficult quiz with questions like *“How many U.S. states border Canada?”* (mean score: 1 out of 10) and they saw quizzes from five who had taken a simple quiz with questions like *“The Bronx is part of what U.S. city?”* (mean score: 8.9 out of 10). These ten quizzes were selected such that the mean score and standard deviation for each type of quiz roughly matched the mean and standard deviation among all quiz-takers in this study from which the quizzes were selected. We divided these 10 quizzes into 2 sets, creating 2 conditions: each of which had 5 easy quizzes and 5 hard quizzes. Within each of these conditions a new condition was created with the order of presentation of the quizzes reversed. Participants were randomly assigned to one of these four conditions.

After seeing a candidate's completed quiz, participants were then reminded of the candidate's score (out of 10), and were told the average score and standard deviation among all 10 test-takers on that quiz. For the first set, the five easy scores participants saw were 9, 9, 9, 7 and 10 ($M = 8.8$, $SD = 1.1$). For the second set, the five easy scores were 10, 8, 8, 9 and 10 ($M = 9.0$, $SD = 1.0$). The five difficult scores participants saw from the first set were 1, 2, 0, 2 and 0 ($M = 1.0$, $SD = 1.0$). For the second set, the five difficult scores were 2, 1, 2, 0 and 0 ($M = 1.0$, $SD = 1.0$). Participants were then asked to rate how knowledgeable about US states and geography they thought each contestant was using a 7-point scale that ran from "*not knowledgeable at all*" to "*very knowledgeable*."

Before making their selections, participants were reminded that each candidate they had seen had either taken a simple or a difficult quiz. Three questions then asked participants to compare the two quizzes on three dimensions:

- 1) The first question asked "*Do you think the two tests were equally good at testing candidates' trivia skills?*" Participants were provided with a response scale with seven asterisks arranged horizontally on the page. The left most asterisk was labeled "*simple is better*" the middle one was labeled "*equal*" and the right most asterisk was labeled "*difficult is better*." For the sake of analyzing the data, responses were coded from 1 to 7, with 1 favoring the simple quiz and 7 favoring the difficult quiz, with 4 indicating neutrality.
- 2) The second question asked "*Do you think the two tests were equally fair measures of ability?*" The endpoints were labeled "*simple is more fair*" and "*difficult is more fair*."

- 3) The third question asked “*Do you think the two tests will be equally good predictors of performance if chosen for the team?*” The endpoints were labeled “*simple is better*” and “*difficult is better.*”

After they had reviewed all ten candidates, participants read, “*After you have spent some time evaluating the previous quizzes, please select which candidates you think will do best on a quiz that was given to all quiz-takers. A copy of this quiz is below. Remember that for each person you select who performs better than average on the quiz at the bottom of this page you will earn \$2 dollars.*” The third test was also a geography test, of intermediate difficulty. After reading these instructions, participants indicated which of the ten candidates they chose.

After they made their selections, their choices were scored and payoffs were computed. After being paid, participants were thanked, debriefed, and dismissed.

Results and Discussion

We computed two averages for ratings of knowledgeability: one for the five contestants whose easy quizzes participants saw, and another for the five contestants whose difficult quizzes participants saw. These averages were then submitted to a (2) (test difficulty) $\times$ 2 (experience) mixed ANOVA with repeated measures on the first factor. The results reveal a significant within-subjects effect of test difficulty, $F(1, 69) = 136, p < .001, \eta^2 = .66$. When participants saw a contestant’s easy quiz, that contestant was rated as significantly more knowledgeable ($M = 5.13, SD = 1.13$) than was the same contestant was rated by participants who had seen his or her difficult quiz ($M = 2.65, SD = 1.01$). The main effect of experience is not significant, $F(1, 69) < 1, p = .59$.

If experience helped participants avoid the correspondence bias, it would have shown up as an experience $\times$ difficulty interaction, wherein experience reduced the effect of difficulty on

rated knowledgeability. This interaction does not attain significance, $F(1, 69) = 3.12, p = .08, \eta^2 = .04$.

Participants were also more likely to pick contestants whose easy quiz scores they had seen when predicting which contestants would score better on the third quiz. Although those who had taken the easy quiz represented 50% of the contestants participants saw, they represented 68% of contestants selected. This 68% is significantly above the 50% we would have expected, had participants perfectly predicted contestants' scores on the third quiz, and only selected those, $t(70) = 7.30, p < .001$. It is also significantly above the 60% we might have expected if participants had been following a justifiable strategy of picking the top two scorers on the easy and difficult quizzes, and then always selecting the next best easy quiz scorer for their fifth pick, $t(70) = 3.35, p = .001$.

In order to compare the results of Study 4 with those of the others, we conducted a binary logistic regression in which selection was the dependent variable. The independent variables in this regression were (1) the quiz score from each contestant that the participant saw and (2) the difficulty of that quiz, as measured by the mean score. The results reveal, consistent with our hypothesis and with the findings of the other studies, that the contestant's actual score was weighted more heavily ($B = 1.61, SE = .13, p < .001$) than was the difficulty of their quiz ($B = -1.37, SE = .12, p < .001$), $\chi^2(1) = 81.10, p < .0001$. In this case, the discounting effect due to quiz ease was 85% the size of the effect of quiz performance.

Specifically, what this means is that going from an average score on the difficult quiz (1.11 out of 10) to an average score on the easy quiz (8.78 out of 10) increases a contestant's probability of being selected from 27% to 70%. This effect is illustrated in Figure 1. In order to construct this graph, we conducted two binary logistic regressions using quiz score performance

to predict the probability of being selected. One regression used easy quiz scores and another used difficult quiz scores. The results show a large effect for quiz difficulty, where easy quiz takers were substantially more likely to be chosen regardless of their performance on the quiz.

When participants were then asked to explicitly compare the virtues of the easy and the difficult quizzes, ratings uniformly favored the difficult test as indicated by the fact that each rating is above the rating scale's midpoint of 4. Participants rated the difficult quiz as a better test of ability than the simple quiz ($M = 4.54$, $SD = 1.95$), $t(70) = 2.32$, $p = .024$. They rated the difficult test as more fair than the simple quiz ($M = 4.44$, $SD = 1.87$), $t(70) = 1.97$, $p = .053$. They also rated the difficult quiz as a better predictor of future performance than the easy quiz ($M = 4.75$, $SD = 1.90$), $t(70) = 3.31$, $p = .001$. It would appear that the only way to reconcile these ratings with participants' systematic preference for takers of the easy quiz is that they believed that the difficult test was better at revealing just how inept the takers of the difficult quiz were.

General Discussion

The results of the four studies we present here are consistent in showing that information about the strength of the situation—in this case, task difficulty—tends not to be used sufficiently to discount information about an individual's performance, even when performance and the situation's influence on it are obvious and are quantified. As a result, students from institutions with lenient grading benefit from their high grades, because their grades are not sufficiently discounted. High grades are easier to achieve and so are less indicative of high performance than are the same grades that were earned from an institution with lower average grades.

It is often the case that structural and situational factors are the most powerful influences on behavior. Within organizations, for example, it is easier to succeed in some jobs than in

others (Pfeffer & Davis-Blake, 1986). Sometimes people will achieve positive outcomes simply because of a beneficent environment. For instance, it is easier to achieve success as a manager when your team is strong than when your team is weak. Likewise, it is easier to obtain a strong education in an excellent private school than in an under-funded public school. And it is easier to achieve high grades at schools where everyone is graded leniently. So it would be a mistake to neglect situational effects on performance, but that is what our data suggest that people do.

The Robust Correspondence Bias

These results contribute to our understanding of the correspondence bias. They suggest that neither underestimating the impact of the situation nor overestimating the impact of behavior are necessary conditions for producing the correspondence bias. In addition, the results suggest that the correspondence bias can persist even when information about both behavior and situation are known with equal clarity and are presented in the same format and modality. This rules out both the main alternative explanations for findings of the correspondence bias, and suggests that the effect is remarkably robust. Perhaps more importantly, the present results afford a useful quantification of the size of the correspondence bias. Its hallmark is that the judgmental weighting attached to the situation is lower than that attached to the behavior, despite the fact that they should be equal. In our results, we find that the situation exerted from 22% to 85% on the influence of performance on attributions of individual ability. Clearly, factors that varied between our studies influenced the size of the correspondence bias. Identifying the moderators of the size of this effect will be an important task for future research on the topic.

Nevertheless, it is worth noting that even being able to make specific estimates for exactly how much the correspondence bias affects judgment represents a departure from prior work on the topic. Like too much of the research in behavioral social science, work on the

correspondence bias has been content to make only directional hypotheses and demonstrate only directional conclusions. In other words, research has concluded that people neglect the strength of situational influences on behavior, *but has failed to tell us how much*. As a result, we are left with a vague directional theories and findings. We can do better. Any construct worth studying is worth measuring properly, and precise measurement can help us go beyond vague directional predictions and make more precise estimates of outcomes and effects.

Contributions to Theory

Comparative judgment. These results contribute to the literature on the psychological process at work in comparative judgment, a literature that stretches across psychology (Windschitl, Kruger, & Simms, 2003), economics (Camerer & Lovallo, 1999), and organizational behavior (Moore, 2004; Moore & Cain, in press). In this paper, we extend previous research by examining judgmental contexts in which decision-makers are comparing outcomes that vary with respect to both nominal performances and their ease. We should also point out that these results are, in a number of ways, more dramatic than the results of previous research showing biases in comparative judgment. Previous results have been strongest when participants themselves are the focus of judgment (Moore & Kim, 2003; Windschitl et al., 2003). Biases in comparative judgment shrink when people are comparing others, and shrink still further when they have perfect information about performance by those they are comparing (Moore, 2007). Biases disappear when comparisons are made on a forced ranking scale (Klar & Giladi, 1997). In this paper, we have shown comparative judgments to be powerfully biased even when people are evaluating others about whom they have complete information, and even when they employ a forced ranking (i.e., admission decisions).

Attribution. These results also contribute to the literature on attribution by demonstrating the persistence of the correspondence bias even in situations where both the behavior and the situation are known with unambiguous certainty. Underestimating the power of the situation and overestimating the distinctiveness of behavior are both rather implausible explanations for the present findings. It is possible but unlikely that our participants were trying to apply the Lewinian equation, but their subtraction skills failed them. Theories reliant on egocentric biases (Kruger, 1999; Windschitl et al., 2003) are not particularly useful for explaining the persistence of the correspondence bias, given that for those making the decisions, the self was not relevant to the comparative judgments we examine. We can also rule out other theories of attribution or comparative judgment based on differentials in the quality of information people have about the target of judgment and the referent (Moore & Small, in press). In our experiments, participants had the same information about all candidates. What theory best explains these results?

Causal Mechanisms. We believe that best explanation for our results is that people habitually take nominal performance (i.e., observable behavior) at face value without adjusting its interpretation based on the situation that gave rise to the performance. The effect we obtain bears striking resemblance to the representativeness heuristic (Kahneman & Tversky, 1973): High grades are usually representative of high ability and future performance. Intuitive judgments are insufficiently sensitive to the reliability or diagnosticity of the evidence, and instead assume that the prior evidence is representative of future promise. People often rely superficially on absolute performance as an indicator of relative performance even when they ought to know better (Baron, 1997; Moore, 2007). While it is clear that people considered the situation to some extent (applicants from schools with lenient grading were less likely to get in, given the same GPA) they were also clearly underweighting it.

The underweighting of task ease was smallest in our fourth study. There at least two obvious reasons that could account for this. First, decisions occurred outside the admissions context, minimizing any possible concern that they could have believed that task ease (lenient grading) was actually associated with higher ability (more selective institutions). Second, participants had a direct stake in making an accurate prediction—they were paid based on the quality of their predictions.

If having money on the line led participants to devote more care and attention to the problem, then it may be that conscious, deliberate effort can help people adjust for situational effects and reduce their susceptibility to this bias. However, actual admissions decisions tend to be made in the presence of a surplus of information and under time pressure. Consistent with research on both the correspondence bias and the representativeness heuristic, we would predict that the effect we have documented would only grow stronger under cognitive load or under time pressure, when people have fewer cognitive resources to devote to it (Gilbert & Jones, 1986; Gilbert, Pelham, & Krull, 1988).

Limitations

One potential objection to external validity of Studies 2 and 3 is that participants were not professional admissions staff. To readers who wonder whether admissions officers could have made better decisions than would undergraduate students, we offer three responses. First, we believe we are studying fundamental features of the psychology of comparative judgment and we are interested in the psychological processes that are common to all types of comparative judgments. As such, these are processes that should be common to people whether they happen to work in admissions departments or not. Second, while it is tempting to believe that the decisions of working professionals will show less bias than those of undergraduate students, the

evidence suggests instead that we should assume the opposite: Unless there is a compelling reason to expect a specific difference between two populations of people we should assume that they are the same, especially with regard to basic psychological processes in judgment and decision making. There is simply too much evidence showing that experienced professionals display the same decision-making biases as do less sophisticated decision-makers (Camerer, 2000; Kahneman, 2003; Kahneman & Tversky, 2000; Lichtenstein & Slovic, 1973). There are, of course, many professions in which expertise really does improve decision-making skill. The hallmark of these situations is that decision-makers obtain clear and timely feedback on their decisions that allow them to correct errors and biases in their judgments (Camerer, 1995; Lichtenstein, Fischhoff, & Phillips, 1982). This cannot be said of admissions decisions, because: (1) selection eliminates the possibility of identifying false negatives (people who were rejected but who would have done well); and (2) outcome data, as measured by the performance and job placement of admitted students, is ambiguous, takes years to collect, and is generally not fed back to admissions offices in any systematic way. Third, the consistency between the results of our laboratory experiments (Studies 2, 3 and 4) and our field study ought to provide some measure of assurance.

Social psychological research has repeatedly shown that people discount situational influences on behavior (Gilbert, 1994; Jones & Davis, 1965; Ross, 1977). Instead, people make inferences about individual differences of skill and temperament based on observed behaviors with insufficient regard to the situational causes that gave rise to those behaviors. However, research in this tradition leaves participants to their own devices to speculate about how much the situation in question actually does influence behavior. When the typical American is compelled to do so, how convincing an essay does he or she write in support of Fidel Castro

(Jones & Harris, 1967)? Exactly how much more knowledgeable does the average quiz-master seem than the average quiz-taker (Ross et al., 1977)? Exactly how much does being late influence seminarians' willingness to stop to help someone in need (Darley & Batson, 1973; Pietromonaco & Nisbett, 1982)? The present research, by contrast, informed participants *exactly* how strong the situation was, thereby making it transparently clear exactly how much performance ought to be discounted. Participants in the fourth study, for instance, knew what the mean scores on both the easy and difficult quizzes were when they were rating applicants. The strong tendency to favor those who had taken the easy quiz attests to the durability of the correspondence bias, even in the presence of full information about the strength of situational influences on behavior.

Conclusion

All four studies supported the hypothesis that people rely heavily on nominal performance (such as GPA) as an indicator of success while failing to sufficiently take into account information about the distributions of performances from which it came. To the extent that admissions officers generally show the same biases we found, graduate programs are collectively choosing to select students who come from undergraduate programs with lenient grading rather than selecting the best students. The consequences could be substantial for both the quality of students selected and the quality of those graduate programs.

The question of whether people—especially decision makers such as admissions officers—can correct for the correspondence bias in judgments of others is fundamental to problems of social inequality and class mobility. A meritocracy depends on being able to identify merit that, in reality, is often clouded by variations in circumstance. Given persistent disparities in the difficulty of the conditions into which Americans are born (Kozol, 1991;

Wilson, 1990), it is essential for colleges and employers to be able to adjust their estimations of ability appropriately based on the ease with which individual promise can result in nominal performance. The results of the present study suggest pessimism—people will too often be judged based on their nominal performances, with insufficient regard to the difficulty of achieving those results.

While universities do prefer to maintain a reputation for high academic standards, the results presented here show that they derive a benefit from having graduates with higher grades: better placements for their alumni. While our studies focused on university admissions decisions, the same effect may also hold for private employers that use grades as a selection criterion. Given the tremendous value of being able to place alumni in better graduate schools and in better jobs, universities cannot be expected to go too far in seeking to curtail grade inflation. For example, universities are highly unlikely to implement meaningful institutional changes such as replacing grades with percentile rankings. Instead, we should expect academic institutions to pay lip service to the importance of high academic standards while at the same time avoiding publicizing average grade distributions and avoiding reporting class rank data on their students.

Can we offer more constructive advice to those in admissions offices, personnel offices, and hiring committees responsible for making such selection decisions? We believe that we can. The advice is consistent with a great deal of other evidence that demonstrates the superiority of actuarial vs. clinical judgment (Dawes, 1972, 1979; Dawes, Faust, & Meehl, 1989; Grove & Meehl, 1996). The advice is that decision makers should not rely exclusively on their unaided intuitive judgments and they should instead obtain the help of a computational decision tool. In this case, what that means is simply that GPA ought to be considered exclusively as a percentile

rank or z -score deviation from the mean at that person's school. Given the power and persistence of the effect we document, the implication seems to be that decision makers should not be allowed to see raw scores or absolute GPA and should only see the standardized score that shows relative performance. Or better yet, each applicant's performance relative to peers ought to be used as an input in a linear model that is used to make admissions or selection decisions.

References

- Baron, J. (1997). Confusion of relative and absolute risk in valuation. *Journal of Risk and Uncertainty*, 14, 301-309.
- Bless, H., Strack, F., & Schwarz, N. (1993). The informative functions of research procedures: Bias and the logic of conversation. *European Journal of Social Psychology*, 23(2), 149-165.
- Camerer, C. F. (1995). Individual decision making. In J. H. Kagel & A. E. Roth (Eds.), *The handbook of experimental economics*. Princeton, NJ: Princeton University Press.
- Camerer, C. F. (2000). Prospect theory in the wild: Evidence from the field. In D. Kahneman & A. Tversky (Eds.), *Choices, values, and frames* (pp. 288-300). New York: Russell Sage Foundation.
- Camerer, C. F., & Lovallo, D. (1999). Overconfidence and excess entry: An experimental approach. *American Economic Review*, 89(1), 306-318.
- Darley, J. M., & Batson, C. D. (1973). "From Jerusalem to Jericho": A study of situational and dispositional variables in helping behavior. *Journal of Personality and Social Psychology*, 27, 100-108.
- Davis-Blake, A., & Pfeffer, J. (1989). Just a mirage: The search for dispositional effects in organizational research. *Academy of Management Review*, 14(3), 385-400.
- Dawes, R. M. (1972). In defense of "bootstrapping." *American Psychologist*, 27(8), 773-774.
- Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, 34(7), 571-582.
- Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, 243(4899), 1668-1674.
- Gilbert, D. T. (1994). Attribution and interpersonal perception. In A. Tesser (Ed.), *Advanced social psychology* (pp. 99-147). New York: McGraw-Hill.
- Gilbert, D. T., & Jones, E. E. (1986). Perceiver-induced constraint: Interpretations of self-generated reality. *Journal of Personality and Social Psychology*, 50(2), 269-280.
- Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, 117(1), 21-38.
- Gilbert, D. T., Pelham, B. W., & Krull, D. S. (1988). On cognitive busyness: When person perceivers meet persons perceived. *Journal of Personality and Social Psychology*, 54(5), 733-740.
- Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law*, 2(2), 293-323.
- Jones, E. E., & Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. *Advances in Experimental Social Psychology*, 1, 219-266.
- Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology*, 3, 1-24.
- Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, 58(9), 697-720.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237-251.
- Kahneman, D., & Tversky, A. (Eds.). (2000). *Choices, values, and frames*. New York: Russell sage Foundation.

- Klar, Y., & Giladi, E. 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(5), 885-901.
- Kozol, J. (1991). *Savage inequalities: Children in America's schools*. New York: Crown Publishing.
- Kruger, J. (1999). Lake Wobegon be gone! The "below-average effect" and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology*, 77(2), 221-232.
- Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art in 1980. In D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306-333). Cambridge, England: Cambridge University Press.
- Lichtenstein, S., & Slovic, P. (1973). Response-induced reversals of preference in gambling: An extended replication in Las Vegas. *Journal of Experimental Psychology*, 101(1), 16-20.
- Martin, S. L., & Klimoski, R. J. (1990). Use of verbal protocols to trace cognitions associated with self- and supervisor evaluations of performance. *Organizational Behavior & Human Decision Processes*, 46(1), 135-154.
- Mitchell, T. R., & Wood, R. E. (1980). Supervisor's responses to subordinate poor performance: A test of an attributional model. *Organizational Behavior and Human Decision Processes*, 25(1), 123-138.
- Moore, D. A. (2004). Myopic prediction, self-destructive secrecy, and the unexpected benefits of revealing final deadlines in negotiation. *Organizational Behavior and Human Decision Processes*, 94(2), 125-139.
- Moore, D. A. (2007). Not so above average after all: When people believe they are worse than average and its implications for theories of bias in social comparison. *Organizational Behavior and Human Decision Processes*, 102(1), 42-58.
- Moore, D. A., & Cain, D. M. (in press). Overconfidence and underconfidence: When and why people underestimate (and overestimate) the competition. *Organizational Behavior & Human Decision Processes*.
- Moore, D. A., & Kim, T. G. (2003). Myopic social prediction and the solo comparison effect. *Journal of Personality and Social Psychology*, 85(6), 1121-1135.
- Moore, D. A., & Small, D. A. (in press). Error and bias in comparative social judgment: On being both better and worse than we think we are. *Journal of Personality and Social Psychology*.
- Nisbett, R. E., & Borgida, E. (1975). Attribution and the psychology of prediction. *Journal of Personality and Social Psychology*, 32(5), 932-943.
- Pfeffer, J. (1977). Effects of an MBA and socioeconomic origins on business school graduates' salaries. *Journal of Applied Psychology*, 62(6), 698-705.
- Pfeffer, J., & Davis-Blake, A. (1986). Administrative succession and organizational performance: How administrator experience mediates the succession effect. *Academy of Management Journal*, 29(1), 72-83.
- Pfeffer, J., & Fong, C. T. (2002). The end of business schools? Less success than meets the eye. *Academy of Management Learning and Education*, 1(1), 78-95.
- Pfeffer, J., & Fong, C. T. (2004). The business school 'business': Some lessons from the US experience. *Journal of Management Studies*, 41(8), 1501-1520.

- Pietromonaco, P. R., & Nisbett, R. E. (1982). Swimming upstream against the fundamental attribution error: Subjects' weak generalizations from the Darley and Batson study. *Social Behavior and Personality*, 10, 1-4.
- Rojstaczer, S. (2003). *Grade inflation at American colleges and universities*. Retrieved October 21, 2003, from <http://www.gradeinflation.com/>
- Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10, pp. 173-220). New York: Academic.
- Ross, L., Amabile, T. M., & Steinmetz, J. L. (1977). Social roles, social control, and biases in social-perception processes. *Journal of Personality and Social Psychology*, 35(7), 485-494.
- Schwarz, N. (1996). *Cognition and communication: Judgmental biases, research methods, and the logic of conversation*. Mahwah, NJ: Erlbaum.
- Staw, B. M., Bell, N. E., & Clausen, J. A. (1986). The Dispositional Approach to Job Attitudes: A Lifetime Longitudinal Test. *Administrative Science Quarterly*, 31(1), 56-68.
- Staw, B. M., Sutton, R. I., & Pelled, L. H. (1994). Employee positive emotion and favorable outcomes at the workplace. *Organization Science*, 5(1), 51-71.
- U.S. News and World Report. (2005). *America's best colleges*. Naperville, IL: Sourcebooks.
- Wilson, W. J. (1990). *The truly disadvantaged: The inner city, the underclass, and public policy*. Chicago: University of Chicago Press.
- Windschitl, P. D., Kruger, J., & Simms, E. (2003). The influence of egocentrism and focalism on people's optimism in competitions: When what affects us equally affects me more. *Journal of Personality and Social Psychology*, 85(3), 389-408.

Footnotes

¹ Assuming similarity across institutions in both quality and in the variance of average students' GPAs across institutions.

² Ideally, we would have obtained data on the average grades at graduation among each candidate's classmates who shared the same major. Since we were unable to obtain such specific historical data, we were forced to use what we could get—usually, average grades among all graduates for a recent year. This introduces at least two problems. First, if average grades have gone up over time, then our measure of comparative performance will underestimate the actual performance of older applicants whose undergraduate GPAs are less inflated. Second, if grades in some majors (such as the natural sciences) are lower than grades in other majors (such as the humanities), then our measure of comparative performance will underestimate the actual performance of graduates of harshly-graded majors. While these are important concerns, we do not believe that they represent confounds that threaten the internal validity of our findings.

³ Of course, one way to decrease the impact of any given independent variable in regression is to include others. However, the non-significance of institution GPA was not a result of the inclusion of eleven other control variables. We conducted another regression that included only institution quality, applicant's GPA, and institution GPA as predictors. The result is that that institution quality ($B = .61, t = 5.33, p < .001$) and applicant's GPA ($B = .49, t = 6.47, p < .001$) remain significant positive predictors of admissions decisions. However, institution average GPA remains insignificant ($B = .02, t = .11, p = .91$).

⁴ Note that in order to reduce the obviousness of our manipulation, both the GPAs of the individual applicants and the average GPAs of their institutions varied slightly around these precise points (within .02). This is the case for Studies 2 and 3.

Table 1

Descriptive statistics and bivariate correlations (Study 1). Correlations in bold are statistically significant ($p < .05$).

	Mean	S.D.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Admission	.50	.50																
2. Male	.71	.45	-.05															
3. U.S. Citizen	.87	.33	.20	0														
4. Age at time of application	27.65	3.26	-.08	.10	-.02													
5. African ancestry	.08	.28	-.08	-.12	.08	.04												
6. Asian ancestry	.19	.39	.05	-.07	.08	-.03	-.14											
7. Hispanic ancestry	.06	.23	-.03	.02	.06	.06	-.07	-.12										
8. Native American	.005	.07	.02	.01	.03	.01	-.02	-.03	-.02									
9. GMAT score	655	79	.36	.19	.04	-.04	-.31	.08	-.15	-.01								
10. Interview score	.04	.76	.34	-.11	.08	.02	.03	.01	-.01	.03	0							
11. Years work history	3.76	2.48	.11	.11	.09	.46	.07	0	.05	-.02	-.05	.05						
12. Undergrad. Institution quality	.55	.35	.19	-.09	.12	-.11	-.09	.13	-.02	-.04	.19	.07	-.07					
13. Private school	.53	.5	.09	-.07	.02	-.08	-.02	-.04	-.03	0	.06	.08	-.07	.40				
14. Tuition	\$20,855	\$10,426	.15	-.10	.06	-.11	-.06	.01	-.02	-.03	.10	.09	-.06	.66	.84			
15. Average Entrance SAT score	1247	137	.17	-.08	.08	-.07	-.11	.11	0	-.03	.20	.08	-.09	.80	.48	.66		
16. Undergraduate GPA	3.32	.42	.19	-.10	-.06	-.12	-.17	-.03	-.05	-.03	.18	.03	-.15	-.13	.01	-.08	-.09	
17. Average GPA at undergrad. institution	3.21	.17	.11	-.07	.11	-.07	-.03	0	-.06	-.01	.16	.10	-.13	.51	.48	.46	.58	.06

Table 2

Unstandardized B weights (and standard errors) from binary logistic regressions predicting applicants' admission.

	Model 1	Model 2
Constant	-18.71** (1.37)	-17.03** (2.52)
Graduate School 1	1.45** (0.17)	1.62** (0.23)
Male	-0.68** (0.15)	-0.8** (0.21)
U.S. Citizen	1.27** (0.21)	1.23** (0.29)
Age	-0.08* (0.03)	-0.11* (0.03)
African ancestry	0.38 (0.27)	0.15 (0.37)
Asian ancestry	-0.03 (0.16)	-0.36 (0.22)
Hispanic ancestry	0.48 (0.29)	0.5 (0.39)
Native American	1.02 (1.01)	1.22 (1.14)
GMAT score	0.02** (0)	0.02** (0)
Interview z-score	1.26** (0.1)	1.24** (0.12)
Years work experience	0.12** (0.04)	0.1* (0.05)
Undergraduate institution quality	0.97** (0.19)	1.22** (0.34)
Applicant's undergraduate GPA	1.38** (0.18)	1.53** (0.24)
Average GPA at undergraduate institution	-	-0.55 (0.6)
Nagelkerke R^2	.51	.53
N	1683	990

* $p < .05$, ** $p < .001$

Table 3

Ratings of undergraduate success, estimated probability of being offered admission, and observed probability of being admitted to graduate school based on undergraduate GPA and average grades at undergraduate institution (Study 2). Standard deviations appear in parentheses.

Institution average GPA	Low $\approx$ 2.4			Medium $\approx$ 3.0			High $\approx$ 3.6		
Individual GPA	.3 below average	About average	.3 above average	.3 below average	About average	.3 above average	.3 below average	About average	.3 above average
Rated prior success (1-7)	2.33 (1.00)	3.42 (0.92)	4.16 (0.96)	3.05 (0.85)	4.25 (0.91)	5.05 (0.78)	3.64 (1.01)	4.49 (1.00)	6.25 (0.75)
Rated probability of acceptance	20% (14%)	36% (17%)	45% (18%)	32% (18%)	47% (19%)	61% (17%)	44% (19%)	54% (20%)	83% (15%)
Actual acceptance rate	2%	7%	56%	7%	50%	96%	30%	61%	94%

Table 4

Ratings of undergraduate success, estimated probability of being offered admission, and observed probability of being admitted to graduate school based on undergraduate GPA and average grades at undergraduate institution (Study 3). Standard deviations appear in parentheses.

Institution average GPA	Low $\approx$ 2.4			Medium $\approx$ 3.0			High $\approx$ 3.6		
Individual GPA	About average	.6 above average	1.2 above average	.6 below average	About average	.6 above average	1.2 below average	.6 below average	About average
Rated prior success (1-7)	3.46 (0.95)	4.78 (0.93)	6.01 (0.69)	3.02 (0.97)	4.64 (0.68)	5.76 (0.65)	2.58 (0.87)	3.99 (0.82)	5.36 (0.93)
Rated probability of acceptance	37% (18%)	58% (19%)	78% (13%)	32% (17%)	57% (14%)	75% (14%)	27% (17%)	50% (16%)	69% (19%)
Actual acceptance rate	4%	60%	98%	0%	56%	96%	0%	13%	76%

Figure 1

Probability of being selected, conditional on quiz difficulty and score relative to others on that quiz (Study 4).